Vol.1 No.15 24 .August 2012

The Monthly Lecture Meeting

第 15 回 月例発表会

第 1 巻 15 号

同志社大学生命医学部

医療情報システム研究室

Published by the Medical Information System Laboratory

OF Doshisha University, Kyotanabe, Japan

Medical Information System Laboratory

The Monthly Lecture Meeting

Contents

MRI によるふくらはぎの脂肪と筋肉の測定

臼井 智美 . . . 1

快・不快画像提示時における脳血流変化の検討

大村 歩 . . . 4

NIRS と EEG による感情画像刺激に対する効果の検討

林 貴之 . . . 7

Smith Waterman 法による類似波形抽出のための再量子化の量子数の検討

福島 亜梨花 . . . 10

SVM と学習データ選択による病理データのクラス分類アルゴリズムの検討

大堀 裕一 . . . 14

角膜内皮細胞のモデル画像におけるパラメータ制御手法

上堀 聖史 . . . 19

角膜内皮細胞画像特徴量抽出システムにおける学習画像作成の検討

布川 将来人 . . . 23

Federation 型データグリッドにおけるデータレプリケーションの有効性の検討

藤井 亮助 . . . 27

MRIによるふくらはぎの脂肪と筋肉の測定

臼井　智美
Usui TOMOMI

背景：スキンケアやヘアケアや体型など美容への関心が高まっており，その中でも女性の脚の形に対する関心は非常に高い．
研究目的：脚のふくらはぎ（腓腹筋）の脂肪や筋肉を数値的に算出することによって，各人に適したダイエット方法を推奨する．
発表の位置づけ：ふくらはぎの形の測定，得られた画像からの脂肪と筋肉の数値的なデータの算出を行う．
方法：MRI によるふくらはぎの撮像，体脂肪計測ソフトウェアによる脂肪と筋肉の数値的な算出を行う．
結果：パラメータを代用したところ，ふくらはぎ全体は撮像できなかったため，撮像方法の改良が必要である．

1　はじめに

近頃では，男女問わずに美容への関心が高まっており，特に，スキンケアやヘアケア，ダイエットに注目が集まっている．その中でもダイエットでは，より詳細な数値情報を得ることで，その効果を把握することができる．本研究では，数値情報を得る手法として MRI(Magnetic Resonance Imaging)[1] に着目した．MRI では筋肉や脂肪のコントラストが，他の CT(Computed Tomography) などの手法に比べると良く，詳細に部位を撮像することができるため，得られた画像から数値を算出することができる．また，ヒップ株式会社の調査[2] では，女性 990 人に「脚に自信があるか」との質問を行った結果，約 68 %の女性が「自信がない」と答えた．このことからも脚のダイエットに対してニーズが高いということが分かり，本稿では脚に着目した．MRI で脚の撮像を行い[1]，得られた画像から脂肪と筋肉を数値的に算出をすることで，適切なダイエットを行うための指標とすることができると考えられる．

本実験では，提案手法の予備検討として，同志社大学の MRI 装置を用いて女性健常者 1 名のふくらはぎの撮像を行った．さらに撮像した結果について体脂肪測定ソフトウェア Fat Biz Calc を用いて，脂肪量の計測を行った．

2　MRI(Magnetic Resonance Imaging)

2.1　概要

強磁場を利用して，体内の水素原子核の分布や状態を撮像する方法である．人体を構成する血液，脂肪などは，水素原子を含んでおり，この水素原子のひとつひとつが微弱な磁気を帯びている．この体内の水素原子核を利用する．MRI は特に血液，臓器といった軟組織のコントラストに優れており，主に頭蓋内疾患や臓器の腫瘍などの撮像に適しているとされている．また，非侵襲であるということも MRI の利点である．

Fig. 1　実験で使用した MRI

2.2　撮像原理

MRI は核磁気共鳴現象を利用して撮像を行う．この共鳴現象で，水素原子核の磁化方向を変動させ，その変動の仕方から T1，T2 といった核磁気共鳴現象に特有のパラメータを，色々な具合に調整して撮影するがことできる．

2.3　核磁気共鳴

核の磁気モーメントを持つ原子は小さな磁石と見なすことができる．この原子に磁場をかけると，原子核の磁気双極子モーメントが磁場ベクトルの周りを一定の周波数（ラーモア周波数）で歳差運動する．この原子核に対してラーモア周波数と同じ周波数で回転する回転磁場をかけると磁場と原子核の間に共鳴が起こる．この共鳴現象が核磁気共鳴と呼ばれる．

2.4　緩和時間

MRI 装置の静磁場に入ると，体内の水素原子の磁石は S 極か N 極のどちらかに向きがそろい，全体として正味の磁化が生まれ，これを縦磁化と呼ぶ．一般的には，その磁化の向きを縦方向（Z 軸）とする．水素原子の共鳴周波数で電磁波パルスを照射すれば，水素原子の軸は，

横に傾き，倒れる．このとき水素原子の軸を電磁波パルスによって傾いた三次元の磁場ベクトルとして考えると，Z 軸上の縦磁化成分と XY 平面上の横磁化成分に分けることができる．そして，水素原子が電磁波，磁場が切られた後で元に戻るとき，Z 軸上の縦磁化が倒れた状態から最初の状態に回復する時間を T1（縦緩和時間），XY 平面上の横磁化がゼロに戻る, 減衰する時間を T2（横緩和時間）と呼んでいる．これらは物質によって異なるため，体内の物質の緩和時間の違いを利用して，目標の物質のコントラストを上げることで，臓器などの軟組織，脂肪，血液などの水分を含むものの画像化を行う．

3　ふくらはぎの脂肪と筋肉の測定

3.1　研究目的

本研究では体型，特に脚部の形に関心を持つ女性を対象とし，ふくらはぎの体型矯正を支援する手法の提案を目的とする．脚の構成にも脂肪が大半を占める人，筋肉が大半を占める人など，外見からだけでは知ることができない．MRI では体内の水素原子を利用しており，脂肪と筋肉では水素原子の密度が異なり，T1 と T2 に違いがある．この違いから，脂肪と筋肉をコントラストによって区別する．得られた画像を基に，脂肪と筋肉の測定を行い，理想的な脚の形との数値的な差を求めることによって，各人に適した腓腹筋の形のダイエットを推奨する．

3.2　技術的要件

本研究では，正確なふくらはぎの形の測定，得られた画像からの脂肪と筋肉の数値的なデータの算出，得られたデータからの意味の解釈と腓腹筋の体系矯正のアドバイスができるようになることが必要である．本実験では，腓腹筋の形の測定，得られた画像からの脂肪と筋肉の数値的なデータの算出の 2 つの実験を行った．

4　実験

4.1　MRI による脚部の撮像

4.1.1　実験目的

MRI を用いて，ふくらはぎの形の測定を行う．本校の MRI にはふくらはぎ用の撮像パラメータ，コイルが無かったため，くるぶし用のパラメータ，全身用コイルで腓腹筋が撮像できるかを確認した．

4.1.2　実験環境

本実験では MRI 装置として日立メディコ製の Echelon Vega(1.5T) を使用した．被験者は女性健常者 1 名．計測箇所は，左脚のふくらはぎにあたる，膝下から足首までの範囲を撮像対象とした．使用したコイルは全身用コイル．撮像パラメータはくるぶし用で実験を行った

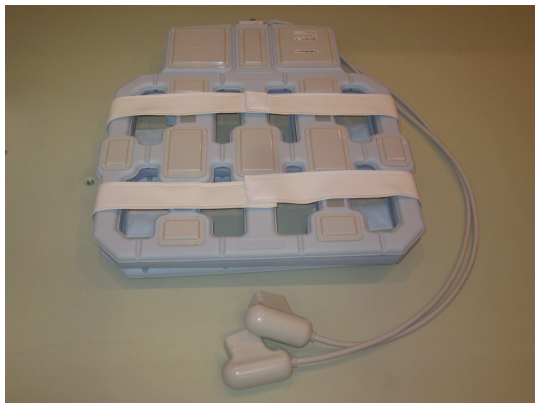


Fig. 2　実験で使用した全身用コイル

Table. 1　パラメータ

スライス厚［mm］	7
スライス数［枚］	17
くり返し時間［msec］	3500
エコータイム［msec］	99
撮像方法	T2 強調画像
撮像断面	axial plane

4.1.3　実験結果

すねから 4.2cm 下の部分，7 枚目のスライスの画像を Fig. 3 に示す．

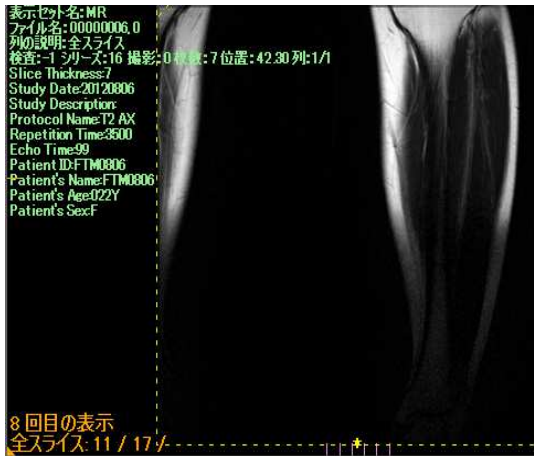


Fig. 3　MRI で撮像された左脚腓腹筋

本実験では，本校の MRI 装置にふくらはぎ用のパラメータがなく，足首用のパラメータを使用したため，撮像範囲が狭く，ふくらはぎから下の部分が撮像できていなかった．このことから，他の撮像方法を考える必要があると考えられる．

4.2　脂肪と筋肉量の測定

4.2.1　実験目的

MRI で撮像された画像からの脂肪と筋肉の数値的なデータが算出できるかを確認した．

4.2.2　実験環境

本実験では，株式会社リジットのメタボリックシンドローム（内臓脂肪症候群）判定のための医用画像計測の体脂肪ソフトウェア Fat Biz Calc を使用し，実験を行った．用いた画像データは，4.1 の実験により得られたデータを使用した．

4.2.3　実験結果

すねから 4.2cm 下の部分，7 枚目のスライスの脂肪測定の画像を示す．なお，ピンクの部分が脂肪である．Fig. 4 ，Fig. 5 に得られた画像を示す．

Fig. 4　Fat Biz Calc で処理された左脚腓腹筋 1

Fig. 5　Fat Biz Calc で処理された左脚腓腹筋 2

Table. 2 に Fat Biz Calc で処理された左脚腓腹筋の計測値を示す．

本実験では，ふくらはぎの下の部分が撮像できていないため，筋肉の部分は背景と同化してしまい，測定を行うことができなかった．脂肪の部分は Fat Biz Calc の性能上，脂肪同士がつながっていないので，2 回に分けて測定を行うことで，撮像できた範囲の測定を行うことができた．脂肪計測ソフトウェアでは撮像断面を axial plane で行うと，輪切りの画像が得られないため，正確な計測ができないことが分かった．これからは coronal plane で測定する必要があると考えられる．

Table. 2　スライスごとの脂肪値

スライス［枚］	脂肪値［$cm\hat{}2$］	スライス［枚］	脂肪値［$cm\hat{}2$］
1	0.0	10	31.5
2	29.4	11	29.5
3	42.8	12	35.6
4	39.3	13	110.5
5	44.0	14	116.3
6	48.1	15	41.5
7	38.6	16	0.0
8	34.3	17	0.0
9	34.5		

5　今後の課題

本研究では，正確なふくらはぎの形の測定，得られた画像からの脂肪と筋肉の数値的なデータの算出，得られたデータからの意味の解釈と腓腹筋の体系矯正のアドバイスができるようになることが必要であるため，本実験では，ふくらはぎの形の測定，得られた画像からの脂肪と筋肉の数値的なデータの算出の 2 つの実験を行った．ふくらはぎの形の測定では，本学校にはふくらはぎ用のパラメータとコイルがなかったので，パラメータは足首用，コイルは全身用を使用し，撮像を行った．しかし，ふくらはぎの下の部分が撮像できていないため，これから，ほかの撮像方法を考える必要があると考えられる．得られた画像からの脂肪と筋肉の数値的なデータの算出では，ふくらはぎの下の部分が撮像できていないことに加え，撮像断面を axial plane で行ったため，脂肪の測定はできたが，筋肉の測定はできなかった．これからは撮像断面を coronal plane で行う必要があると考えられる．また，理想的な足の形との数値的な差を求めるために，標準化を行う必要があると考えられる．

参考文献

1) 株式会社　日立メディコ，日立 MR　イメージング装置　Echelon Vega 取扱説明書，2008
2) ヒップ株式会社，日本・中国・韓国の働く女性を対象に行った「脚」に関する意識調査，http://www.pip-club.com/enjoy/bikyaku/，2012
3) Scott A. Huettel ，Functional Magnetic Resonance Imaging,Sinauer Associaes Inc，2009

快・不快画像提示時における脳血流変化の検討

大村 歩
Ayumi OHMURA

背景：近年，少子高齢化に伴い，要介護者の人口が増加しており，彼らのコミュニケーションを支援する必要がある．
研究目的：コミュニケーションをより円滑に行うことができるような新しいツールを開発する．
発表の位置づけ：感情表現に関係する快・不快における脳血流変化を調査する．
方法：MRI 装置を用いて快・不快の視覚画像を提示する実験を男性健常者 1 名に行った．
結果：視覚情報処理において心象に関与している楔前部に賦活が見られた．

1　はじめに

近年，少子高齢化が進み，老年人口が増加しており，それに伴い，介護を必要とする人口も増加している．要介護者は感情表現が乏しくなることも多く，コミュニケーションをより円滑に行うことができるツールが必要であると考えられる．本稿では，感情表現に関係する快・不快に関する視覚実験を MRI（magnetic resonance imaging）装置を用いて行う．本実験では MRI 装置で撮像する際の動作手順，パラメータ設定および撮像データの処理技術を習得し，感情表現に関係する快・不快における脳血流変化を調査する．

2　実験

2.1　実験目的

本実験では，快・不快な画像を見た際の脳の活性部位を確認する．そのために，先行研究 [1] に基づいて Pleasant（快画像），Unpleasant（不快画像），Neutral（中性画像）を提示し，fMRI（functional magnetic resonance imaging）を用いて脳血流変化を計測する．

2.2　検討項目

先行研究では，Pleasant 提示時は楔前部，下側頭皮質に活性がみられ，Unpleasant 提示時には楔前部，下側頭皮質，外側後頭部に活性がみられた [1]．また，快・不快のどちらにも偏らない Neutral 提示時の賦活領域は下側頭皮質であった．Pleasant および Unpleasant 提示時に賦活する楔前部は心象に関与しているといわれている部位 [2] で，共通して賦活した下側頭皮質は視覚情報の特徴を認識する部位である．外側後頭部は視覚情報の形態の認知に関係している．賦活領域について Fig. 1 に示す．

本実験でも目標の領域，特に楔前部での活性がみられると予想し，後頭極より 10 cm の領域について撮像および検討を行った．

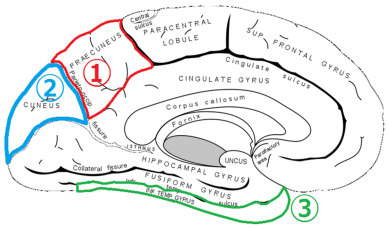


Fig. 1　賦活領域

2.3　被験者

被験者は男性健常者 1 名（24 歳，右利き）を対象とした．

2.4　実験設計

実験に使用する刺激画像は，先行研究で用いられた IAPS（International Affective Picture System）のデータセット [3] を使用した．

本実験の流れを Fig. 2 に示す．最初のレスト 30 秒間は「＋」マークを注視する．その後，刺激画像提示（6 秒間）を Pleasant，Unpleasant，Neutral の順番で提示する．なお，刺激提示と刺激提示の間には 12 秒間のレストを挟む．3 種類の画像提示を 1 ブロックし，このブロックを 10 回繰り返す．これを 1 セッションとし，レストを 1 分挟んで，2 セッション行う．

使用した画像の番号を Table. 1 に示す．

2.5　実験環境

本実験で使用した MRI 装置を Fig. 3 に示す．MRI 装置は日立メディコ製の ECHELON Vega（1.5 T）を使用した．実験を行った際の温度は 23 ℃であった．

実験では，まず頭部位置決め画像を撮像し，機能画像を撮像する．その後，構造画像を撮像する．機能画像撮像のパラメータを Table. 2 に示す．

2.6　データ処理方法

データ処理は SPM8[4] を用いて以下の手順で行った．

1. Realign：体動データの補正を行う．

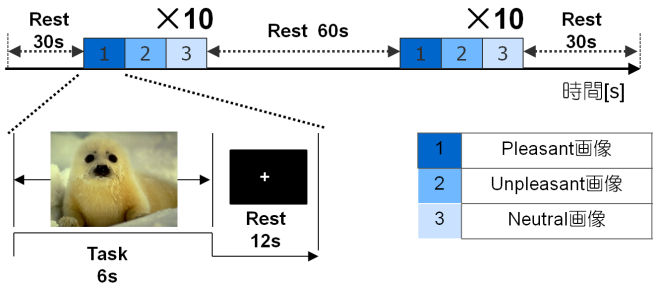


Fig. 2　実験設計

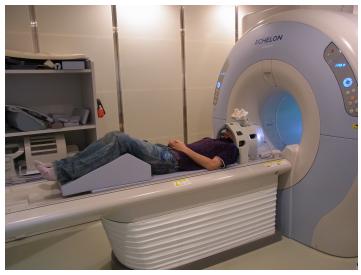


Fig. 3　MRI 装置（ECHELON Vega 1.5T，日立メディコ）

2. Slice timing：各スキャンでの撮像タイミングを合わせる．
3. Smoothing：データの平滑化を行う．HWHM は 8 mm とした
4. Coresister：被験者の機能画像と構造画像を合わせる．
5. Normalize：MNI 標準脳座標系を用いて標準化を行う．
6. t 検定：有意水準 0.000001 で検定を行う．

3　実験結果

Pleasant，Unpleasant，Neutral 提示時の脳の賦活領域をそれぞれ Fig. 4〜Fig. 6 に示す．

1〜3 の番号は Table. 3 の部位に対応するしている．

Pleasant 提示時は楔部，楔前部，下側頭皮質で活性がみられ，Unpleasant の時は楔前部，下側頭皮質で活性がみられた．Neutral 提示時は楔部での活性がみられた．

Table. 1　使用画像

画像種別	番号
Pleasant	4611, 4641, 4658, 4659, 4666, 4676, 4677, 4680, 4681, 4690, 1440, 1460, 1463, 1530, 1540, 1590, 1610, 1710, 1750, 1920
Unpleasant	1010, 1019, 1050, 1052, 1090, 1110, 1111, 1113, 1114, 1120, 3000, 3051, 1060, 3068, 3069, 3071, 3100, 3101, 3266, 3400
Neutral	2191, 2214, 2215, 2372, 2383, 2393, 2394, 2480, 2595, 7550, 5740, 7036, 7041, 7050, 7100, 7130, 7161, 7224, 7234, 7500

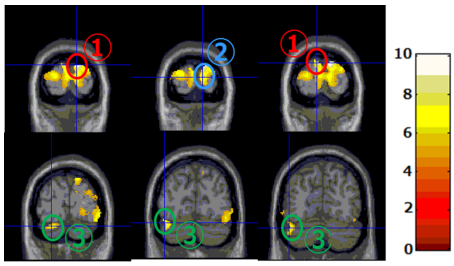


Fig. 4　Pleasant 時の実験結果

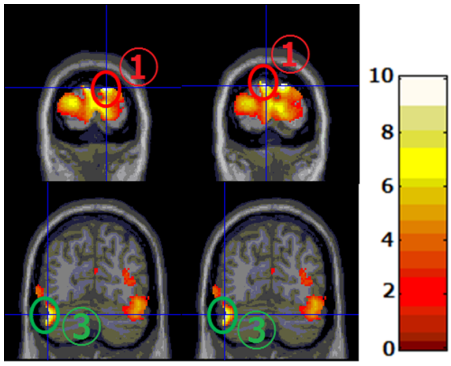


Fig. 5　Unpleasant 時の実験結果

4　考察

実験結果と先行研究での結果を Table. 4 に示す．

先行研究の結果と異なる部分は Pleasant，Neutral 提示時に楔部に活性が見られた点である．楔部は視覚情報の入力を受けている領域であるため，活性したと考えられる．また，本実験では Unpleasant 時に外側後頭部に活性が見られなかった．活性度合の閾値を下げてみたところ，左脳領域での外側後頭部では活性が見られたため，

Table. 2　使用パラメータ

スライス方向	Coronal
シーケンス	GE EPI
画像	T2*強調画像
FOV［mm］	192
TR［ms］	50
TE［ms］	40
スライス厚［mm］	5
スライス枚数	20

Table. 3　賦活領域

1	楔前部
2	楔部
3	下側頭皮質

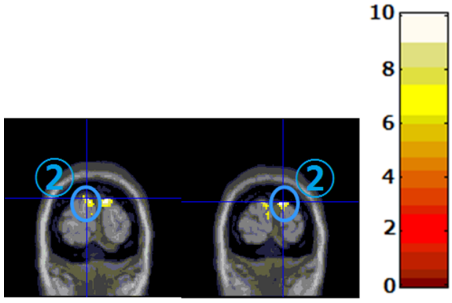


Fig. 6　Neutral 時の実験結果

視覚情報の形態の認知は行えていたと考えられる．情動刺激である Pleasant，Unpleasant 画像提示時に楔前部で活性しており，快・不快に関するデータはとれたと考えられる．

参考文献

1) Dean Sabatinelli, Lang PJ, Andreas Keil, Bradley MM. Emotional perception: Correlation of functional mri and event-related potentials. Technical report, 2006.

2) Fletcher PC, Frith CD, Baker SC et al. The mind ' s eye-precuneus activation in memory-related imagency. Technical report, 1995.

3) Lang PJ, Bradley MM, Cuthbert BN. International affective picture system(iaps): technical manual and affective ratings. Technical report, 1999.

4) Friston KJ, Ashburner JT, Stefan Kiebel, Thomas Nichols, Penny WD. *Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Academic Press Inc, 2006.

Table. 4　先行研究の結果と本実験の結果

	先行研究	実験結果
Pleasant	楔前部，下側頭皮質	楔部，楔前部，下側頭皮質
Unplesant	楔前部，下側頭皮質，外側後頭部	楔前部，下側頭皮質
Neutral	下側頭皮質	楔部

NIRS と EEG による感情画像刺激に対する効果の検討

林　貴之
Takayuki HAYASHI

背景：ALS(筋萎縮性側索硬化症) のような感情を適切に表現できない患者の増加
研究目的：NIRS と EEG を用いた快・不快認識システムの開発
発表の位置づけ：快・不快を想起した際の活性部位の検討
方法：NIRS と EEG を用いて，感情刺激 (快・不快) を行った際の血流と脳波を計測
結果：先行研究通り脳波の左右差を確認した．また，不快感情は両側前頭部が関係しており，不快度が高いほど脳血流が変化した．

1　はじめに

近年，ALS(筋萎縮性側索硬化症) のような感情を適切に表現できない患者が増加している．ALS は，運動神経が障害を受け筋肉が萎縮していく進行性の神経難病である．病気が進行するにしたがって手足をはじめ体の自由がきかなくなるが，頭脳は正常のままである．このような患者が感情を表現するために，快・不快認識システムの開発が望まれる．

本研究では，非侵襲生体計測装置である NIRS (Near Infrared Spectroscopy) と EEG (Electroencephalograph) を用いて快・不快を認識できるシステムの開発を目指す．NIRS の特徴として，体動に強いが，脳表面の信号しか計測することができないという点が挙げられる．一方，EEG は体動などのアーチファクトに弱いが，脳深部の信号まで計測可能である．NIRS-EEG を同時に計測することで，お互いの装置の計測における利点を活用することができ，より正確な脳信号を抽出できると考えられる．

本研究では，その前段階として快・不快時の活性部位について先行研究 [1) 2)] を参考に検討した．

2　実験

2.1　実験目的

本実験の目的は，先行研究を参考に感情画像 (快・不快) を呈示した際の脳波と血流を計測し，快・不快における計測データの差異について検討することである [1) 2)]．また，感情度と血流量が関係性について調査することである．

2.2　被験者と実験環境

被験者は，22 ± 0.8 歳の健康な男性 5 人とし，実験の室温は 24.2 ± 1.0 ℃ の環境下で，15〜18 時の時間帯に実験を行った．NIRS 装置として ETG-7100 (Hitachi medical Co.,Japan) を使用し，計測部位は，前頭部 22CH とし，プローブ設置は脳波電極配置の国際基準である国際 10-20 法に従った．また，EEG 装置として Polymate AP-100 (TEAC Corporation) を使用した．計測部位は，NIRS と同様に国際 10-20 法に従い，基準電極を Cz にし，F3, F4 とした．

2.3　実験設計

本実験では，感情に関する研究で刺激として多くの心理実験で使用されている IAPS(International Affective Picture System)[3)] を刺激として用いた．画像を液晶ディスプレイに呈示し実験を行った．実験に使用した 18 枚の画像番号は，快画像：1710,4599,4660,5202,5621, 5700,8030,8200,8470 不快画像:1300,3000,3060,3102,3150, 3170,3180,6313,6570 である．Fig. 1 に，本実験のタスクシーケンスを示す．

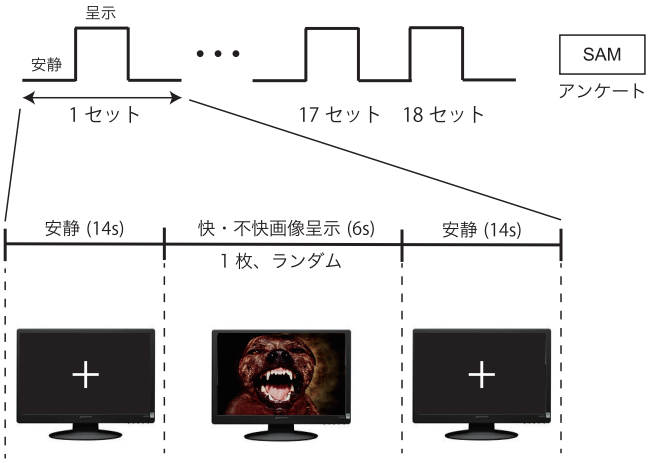


Fig. 1　タスクシーケンス

実験設計は先行研究 [1)] を参考に，ブロックデザインで設計した．以下に，詳しく実験内容について述べる．

1. 安静：14 秒間画面に表示された十字を注視する．
2. 快・不快画像呈示：6 秒間で 1 枚，ランダムに快画像か不快画像のどちらかが呈示される．
3. 安静：14 秒間画面に表示された十字を注視する．

3　検討方法

3.1　NIRS の検討

本実験では，Oxy-Hb(Oxy-hemoglobin) が脳の神経活動に伴い増加するとされているため，Oxy-Hb 濃度変化を用いて検討を行った [4]．脳血流変化の検討部位は，感情が生成される扁桃体からの情報を受けている前頭部に注目し，中でも先行研究で感情画像呈示時に血流変化が見られた左前頭部 (CH1,5,6,10) および右前頭部 (CH4,9,8,13) について検討を行った．また，感情度合いと血流に関係性があるのかを明らかにするために，活性部位を画像別で検討した．

3.2　EEG の検討

脳波を用いた感情についての研究は，広く行われている．なかでも，Davidson らの研究では，大脳の非対称性 (Cerebral Asymmetry) と感情の関連について仮説を立てている [5]．その仮説によると，不快感情は右前頭部に関係し，快情動は左前頭部が関係することが報告されている．この左右差は，脳波でも確認されている．脳波を周波数帯域ごとに解析を行うと 4 つの δ, θ, α, β 波に分類できる．中でも，α 波 (8〜13Hz) は，リラックス時に計測され，この α 波が減衰すると脳の賦活を表す．この α 波帯域のパワー値に注目して左右差の確認が行われている．Davidson らの研究によると，感情的な音楽刺激を行った際に α 帯域パワー値の左右差が確認されている [2]．また，本実験で用いた IAPS を刺激とした研究でも左右差が確認されている [6]．本実験では，EEG で脳波を計測し，α 帯域パワー値の左右差の検討を行った．

4　実験結果

4.1　実験結果 (EEG)

F3 および F4 から計測された脳波に対して，FFT (Fast Fourier Transform) を行い周波数変換し，α 帯域である 8〜13(Hz) の周波数スペクトルを積分し，左右前頭部で比較を行った．比較を行う際に，脳波は個人間で大きく差が見られるため，対数をとり相対値で比較を行った．Fig. 2 に被験者 5 人の F3(左) と F4(右) における α 帯域パワー値を示す．

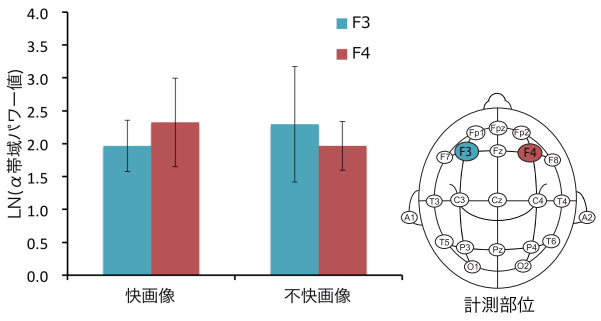


Fig. 2　α 帯域パワー値の左右比較

Fig. 2 より，快画像呈示時に，F4 に比べて F3 の α 帯域パワー値が相対的に減少した．また，不快画像呈示時に，F3 と比べて F4 が相対的に減少した．これは，Davidson らの研究 [2] の結果を支持するものとなった．

4.2　実験結果 (NIRS)

快・不快画像呈示時の活性部位を抽出する為に，各被験者の各チャンネルごとに有意水準 5% で Dunnett 検定 (時間 (n=6) ×試行 (n=9)) を行った．今回の実験では，コントロール群として画像呈示後 1 秒の Oxy-Hb 濃度変化の平均値，処理群として 2,3,4,5,6 秒の平均値のデータを用いた．その一例として，Fig. 3 に被験者 C の不快画像呈示時の CH5 における脳血流変化を示す．

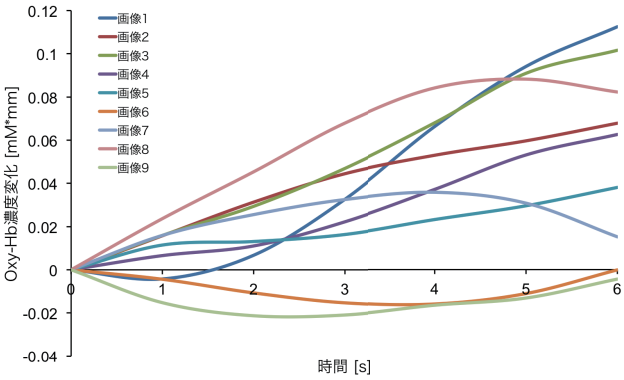


Fig. 3　不快画像呈示時の Oxy-Hb 変化 (被験者 C)

Fig. 3 より，検定の結果 2 秒以降に有意な増加が見られた．この処理を各被験者の各チャンネルに行った結果を Table.1 に示す．

Table. 1　Dunnett 検定により有意な差がみられた CH

	快		不快	
被験者	有意に増加	有意に減少	有意に増加	有意に減少
A	-	14	1/5	14
B	-	14	1/9	-
C	-	-	1/5/9	14
D	-	-	1/5/9	-
E	-	14	5	-

本実験では，被験者 5 人中 3 人以上有意な変化が見られたチャンネルを，変化が見られた部位とした．Fig. 4 に，脳血流変化がみられた部位を示す．

Fig. 4 より，不快画像呈示時において腹外側前頭前皮質 (VLPFC : ventrolateral prefrontal cortex) に有意な血流増加がみられた．これは，先行研究と同じ部位の活性である [1]．快画像呈示時では，有意な血流増加は確認できなかったが，左前頭部で血流減少が確認された．

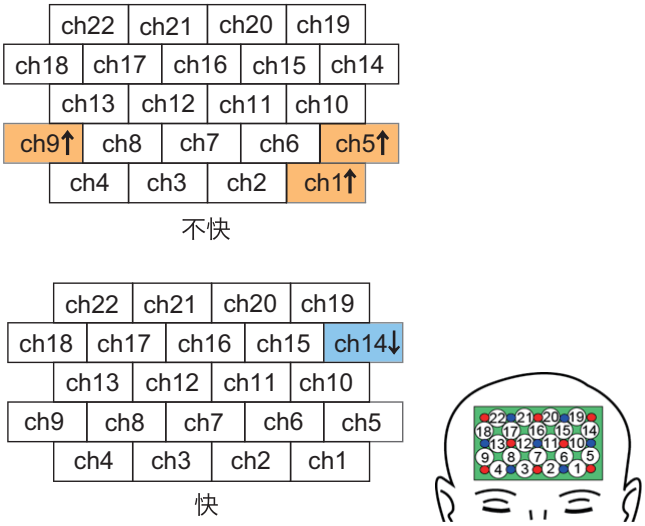


Fig. 4　有意な変化がみられた CH

さらに，不快の度合いによって血流変化に差が見られるのかを検討するために，不快画像呈示時に有意に増加した CH5 について，画像の valence 値によって血流の変化に差異があるかの検討を行った．valence 値とは，9 段階の快・不快度であり，IAPS の各画像に対して数値がつけられている．値が大きい方が快，小さい方が不快な画像である．

Fig. 5 に，被験者 5 名が最も不快な画像 (valance rate 1.69) とあまり不快でない画像 (valance rate 2.59) を見た際の脳血流変化を示す．

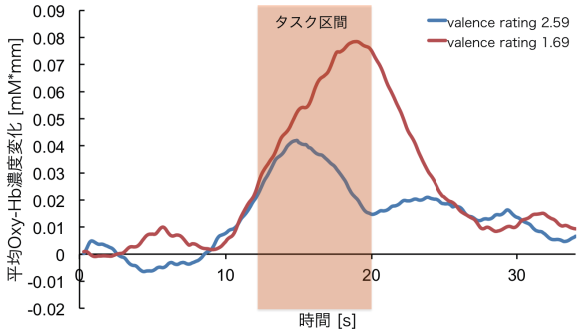


Fig. 5　valence 値による比較

Fig. 5 より，不快度の高い画像を呈示した際に脳血量が変化した．この結果より，不快度の度合いによって，血流量が異なることが示唆された．

5　考察

不快画像時に VLPFC が活性した．VLPFC は注意に関する脳部位であるため，不快画像による注意による活性であると考えられる．近年の研究においても VLPFC が注意に関与していることが報告されている [7]．また，Fig. 5 より，不快度が高いほど血流が増加した．この結果から，不快度が高い画像の方が，画像に対して注意しており，より活性したと考えられる．

快画像呈示時では有意な血流増加がみられなかったが，左前頭部で有意な血流減少がみられた．これは，PET を用いた先行研究と同じ結果を示している．この論文では，より快なほど左前頭部で血流減少が報告されている [8]．

これらの結果より，快・不快の時では，両側前頭部に有意な変化がみられたため，両側前頭部に注目して快・不快認識を行うべきであることが示唆された．

6　まとめ

本稿では，IAPS を用いて感情刺激を行った際の脳の変化を調査した．被験者 5 名に対し実験を行い，快・不快刺激による活性部位を明らかにした．快感情では，左前頭部が有意に減少し，不快時には，両側前頭部が有意に増加した．有意に変化した CH を詳しくみると，感情度合いによって脳血流が変化し，より不快であるほど血流の増加傾向がみられた．これらの実験結果より，感情に関連した前頭部の変化を NIRS と EEG を用いて抽出することができることが示唆された．

参考文献

1) Yoko Hoshi, Jinghua Huang, Shunji Kohri, Yoshinobu Iguchi, Masayuki Naya, Takahiro Okamoto, and Shuji Ono. Recognition of human emotions from cerebral blood flow changes in the frontal region: a study with event-related near-infrared spectroscopy. *J Neuroimaging*, Vol. 21, No. 2, pp. e94–e101, 2011.

2) SCHMIDT L. A. Frontal brain electrical activity (eeg) distinguishes valence and intensity of musical emotions. *Cognition and Emotion*, Vol. 15, No. 4, pp. 487–500, 2001.

3) P. J. Lang, M. M. Bradley, and B. N. Cuthbert. International affective picture system (iaps): Technical manual and affective ratings. Technical report, University of Florida, 1999.

4) Fox PT and Raichle ME. Focal physiological uncoupling of cerebral blood flow and oxidative metabolism during somatosensory stimulation in human subjects. *Proc Natl Acad Sci USA*, Vol. 83, pp. 1140–1144, 1986.

5) Richard J. Davidson, Paul Ekman, Clifford D. Saron, Joseph A. Senulis, and Wallace V. Friesen. Approach-Withdrawal and cerebral asymmetry: Emotional expression and brain physiology i. *Journal of Personality and Social Psychology*, Vol. 58, No. 2, pp. 330–341, 1990.

6) Rene J. Huster, Stephan Stevens, Alexander L. Gerlach, and Fred Rist. A spectralanalytic approach to emotional responses evoked through picture presentation. *International Journal of Psychophysiology*, Vol. 72, No. 2, pp. 212–216, May 2009.

7) Hampshire, Adam, Owen, and M. Adrian. Fractionating attentional control using Event-Related fMRI. *Cerebral Cortex*, Vol. 16, No. 12, pp. 1679–1689, December 2006.

8) Parkekh PI George MS, Ketter TA. Brain activity during transient sadness and happiness in healthy women. *Am J Psychiatry*, pp. 341–351, 1995.

Smith Waterman 法による類似波形抽出のための再量子化の量子数の検討

福島　亜梨花
Arika FUKUSHIMA

背景：多くの時系列データを検出器毎に観察し，特徴的な波形を抽出することは，解析者に大きな負担を強いる
研究目的：時系列データの類似部分を自動的，かつ高速に抽出することで解析者の負担を少なくする
発表の位置づけ：提案手法におけるパラメータの設定がどのような影響を及ぼすのかの検討を行う
方法：Smith Waterman 法による fNIRS 時系列データ類似波形抽出アルゴリズム
結果：データによってパラメータがどのような影響を及ぼすのかが異なるので，データ毎に設定を変える必要がある

1　はじめに

近年，脳機能イメージング装置として，非侵襲で，被験者の自由度が高い functional Near-infrared Spectroscopy (fNIRS) 装置が注目を浴びている [1]．fNIRS 装置の 1 つである ETG-7100(日立メディコ製，日本) は 120 個の検出器を有し，その 1 つ 1 つの検出器から脳血流量の時系列データが得られる．しかし，このような多くの生体情報を検出器毎に観察し，特徴的な波形を抽出することは，解析者に大きな負担を強いることが考えられる．そこで，西井ら [2] は DNA の局所探索法であるホモロジー検索手法として用いられる Smith Waterman 法 [3] を用いた fNIRS 時系列データの類似波形抽出による特徴的波形抽出を行った．本来，Smith Waterman 法は DNA の塩基である「A(アデニン)」,「T(チミン)」,「G(グアニン)」,「T(チミン)」の 4 種類の文字からなる文字列の類似部分を高速に，かつ，高精度に抽出するアルゴリズムである．そこで，西井らのアルゴリズムでは Smith Waterman 法を fNIRS 時系列データに適用するために fNIRS データを 5 または 10 の量子数で再量子化し，文字列に置換している．しかし，5 または 10 の量子数では目視確認だが類似波形以外の波形も抽出しているという問題点がある．そこで，本研究では，類似波形抽出に最適な量子数を求め，Smith Waterman 法による類似波形抽出の精度向上を目的とする．

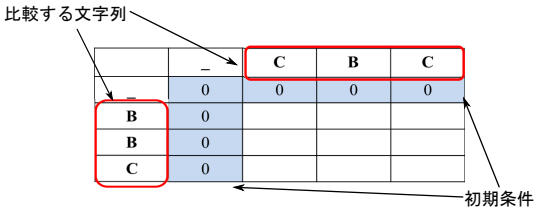
2　Smith Waterman 法による類似波形抽出アルゴリズム

西井ら [2] が提案している類似波形抽出アルゴリズムは以下である．

Step 1 fNIRS データの再量子化 (文字列化)
Waterman 法は類似文字列抽出アルゴリズムのため，数値データである fNIRS 時系列データを文字列に置き換える必要がある．そこで，まず fNIRS データを適当な区間で分割し，文字に置き換える．西井らの先行研究では，量子数 (文字数) を 5 または 10 に設定している．区間分割法として，区間幅が等間隔な等間隔領域分割，または，区間分割するデータを正規分布と仮定し分割する Symbolic Aggregation approXimation(SAX) 法 [4] を用いている．

Step 2 Smith Waterman 法による類似部分抽出
Smith ら [3] が提案している Smith Waterman 法は，進化，系統分類の解析，タンパク質の機能解析などを目的とした配列解析の最も基本的な手法の一つであるホモロジー検索の一種である．Smith Waterman 法は他のホモロジー検索である FAST All(FASTA) や Basic Local Alignment Search Tool (BLAST)[5] と比較し，高精度かつ高速で類似部分を抽出できる．2 つの文字列,「C，B，C」と「B，B，C」を例にして，Smith Waterman 法のアルゴリズムを以下に示す．

1. スコアテーブルの計算
Smith Waterman 法は 2 つの文字列を比較する際，Fig. 1 のようなスコアテーブルを必要とする．Fig. 1 の 1 つ 1 つのセルスコアを式 (1)，() に従い計算していく．なお，式 (1)，() の mismatch は文字列の不一致部分に，gap は文字列中にスペースを発生させることによって類似とみなせる文字列抽出に関わるパラメータである.「C，B，C」と「B，B，C」の例では，ホモロジー検索の際に最も良く用いられる $match = 1$，$mismatch = -1$，



	_	C	B	C
_	0	0	0	0
B	0			
B	0			
C	0			

Fig. 1　スコアテーブル

	_	C	B	C
_	0	0	0	0
B	0	0	1	0
B	0	0	1	0
C	0	1	0	2

最大スコア

Fig. 2　トレースバック

$gap = -1$ で計算する．

$$SW(y,x) = max \begin{cases} SW(y-1, x-1) + match \\ SW(y-1, x) + gap \\ SW(y, x-1) + gap \\ 0 \end{cases} \quad (1)$$

$$SW(y,x) = max \begin{cases} SW(y-1, x-1) + mismatch \\ SW(y-1, x) + gap \\ SW(y, x-1) + gap \\ 0 \end{cases} \quad (2)$$

2. トレースバック
計算した全スコアから最大値を求める．類似部分を求めるために最大スコアから「0」までどのようなスコアを通ってきたかを逆に辿っていく．Fig. 2 のように最大スコアである「2」は左上のスコア「1」を使って求められている．その「1」も左上のスコア「0」を使って求められている．よって，類似部分は「2」と「1」のスコア計算の際に使用した文字列「B，C」と「B，C」になる．

3　実験

実験内容を以下に示す．

3.1　実験目的

本実験では，西井らの提案した Smith Waterman 法による類似波形抽出アルゴリズムの精度の向上のため，量子数が類似部分抽出にどのような影響を及ぼしているのかを検討する．

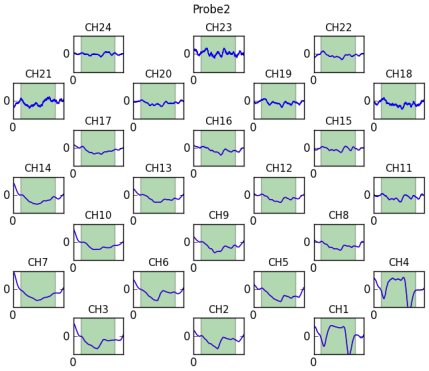


Fig. 3　fNIRS 時系列データ

3.2　対象データ

本実験の対象データはブロックデザインを用いた GO/NOGO Task 実験の fNIRS の酸素化ヘモグロビン濃度データ (Fig. 3) のデータセット A，B，C の 3 セット使用した．なお，測定機器として ETG-7100 を使用した．データセットの詳細を Table. 1 に示す．

Table. 1　fNIRS 時系列データ

パラメータ	値
測定部位	左側頭部
チャネル (検出器) 数	24
ウェイト時間 [s]	30.0
タスク時間 [s]	120.0
レスト時間 [s]	30.0
キャリブレーション時間 [s]	5.0
サンプリング周波数 [Hz]	10.0

3.3　実験設定

Smith Waterman 法による類似波形抽出アルゴリズムに関してのパラメータを以下に示す．

3.3.1　量子数

本実験では，先行研究で用いられている量子数 5，10 とスタージェスの公式 [6] である式 (3) で算出された 3 つで再量子化を行った．ただし，n はデータ数である．スタージェスの公式は，ヒストグラムを作成する際に最も良く用いられる階級数の公式である．本実験では，データの特徴量を表現する手法の一種であるヒストグラムの階級数を用いることによってデータの特徴を踏まえた量子数を実現できるのではないかと考えた．

$$\log_2 n + 1 \quad (3)$$

3.3.2　Smith Waterman 法

先行研究では，$match = 1$，$mismatch = -1$，$gap = -1$ を使用していた．本実験では，文字列が完全一致の類

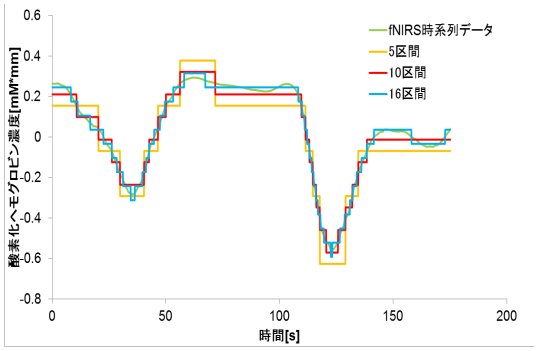


Fig. 4　再量子化

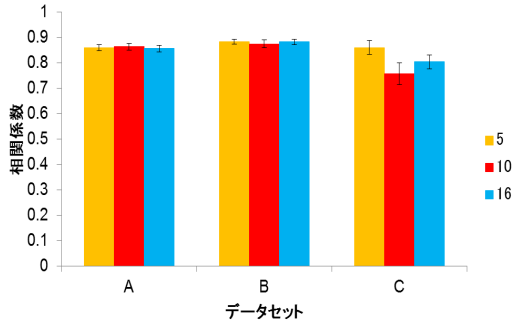


Fig. 5　最大の相関係数

似部分のみを抽出するために *match* を 1 に，*mismatch*，*gap* を比較するデータ長の負数に設定した．

3.3.3　類似度評価

本実験では，類似度の評価として相関係数を用いた [7]．相関係数は式 (4) で表される．相関係数が 1 に近いほど 2 つの波形は類似している．また，絶対値が 0 に近いほど関連性のない波形を抽出している．

$$\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2}} \tag{4}$$

3.4　実験結果

本実験で使用した fNIRS 時系列データのデータ数は全てのデータセットにおいて同様の条件であるので，Table. 1 より 42024 である．これを式 (3) に当てはめると 16 になる．以上より，量子数 5，10，16 で以降は検討していく．量子数 5，10，16 で再量子化したデータセット A の 1 つのチャネルの fNIRS 時系列データを Fig. 4 をに示す．

3.4.1　相関係数

量子数 5，10，16 で再量子化し，Smith Waterman 法を用いて抽出した類似部分の相関係数を求めた．ただし，比較するチャネルの組み合わせにより，Smith Waterman 法の最大スコアが複数あった際は複数の類似部分を抽出している．本実験では，基準のチャネルの fNIRS 時系列データとの類似部分をその他のチャネルの fNIRS 時系列データから抽出することを想定している．そのため，基準のチャネルに対し，最大の相関係数を Fig. 5 に，相関係数の絶対値が 0 最も近い相関係数を Fig. 7 に示した．

相関係数の最大値では，データセット A，B において量子数 5，10，16 の間であまり差が見られなかった．しかし，データセット C においては量子数 10 の際に他の量子数と比較し，差が見られた．以上より，データによって量子数が与える影響が異なることが分かった．

更に量子数が相関係数の最大値に与える影響を調査するためにあまり差が見られなかったデータセット A，B に対し，量子数を 5 から 95 まで 1 ずつ増やし類似波形を抽出し，相関係数の最大値を求めた (Fig. 6)．最大値をとる量子数は，データセット A，B において，51 と 37 であり，最小値をとる量子数は，95 と 7 であることがわかった．Fig. 6 より，最大の相関係数はどのような量子数においても 0.9 から 1 の間であることがわかる．

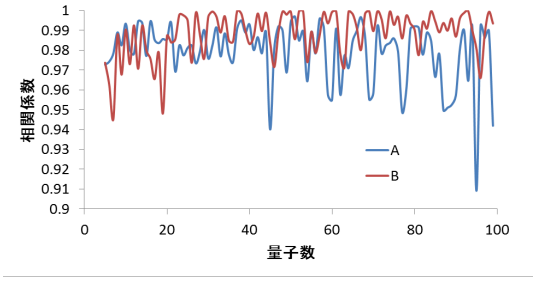


Fig. 6　最大の相関係数と量子数

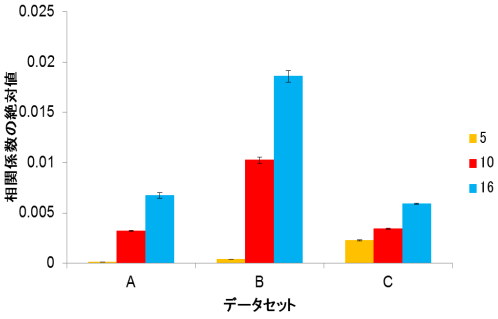


Fig. 7　0 に近い相関係数

0 に最も近い相関係数の絶対値においては，どのデータセットにおいても量子数が 5，10，16 と多くなるにつれ，0 に最も近い相関係数の絶対値が大きくなることがわかった．

3.5　考察

本実験では，最も類似した波形の評価を最大の相関係数を求めて行った．その結果，量子数の影響を受けるが相関係数が 0.9 から 1 の間であり，どのような量子数を設定しても最も類似した波形においては抽出できている

のではないかと考えた．また，最も無関係な波形の評価に関しては，量子数が大きくなるにつれ，0 に最も近い相関係数が大きくなっていった．以上のことから，量子数を大きくするにつれ，無相関である波形を抽出を防ぐことができる可能性があることがわかった．以上より，量子数は類似波形抽出に大きな影響を及ぼすが，どのような影響を及ぼすかはデータによって異なると考えられる．

4 まとめ

本稿では，Smith Waterman 法による fNIRS 時系列データの類似波形抽出アルゴリズムの精度の向上を目的とし，再量子化の際の量子数による影響の検討を行った．最も類似した波形の評価では，どのような量子数を設定しても最も類似した波形に関しては精度を保って抽出できるが，最も無関係な波形の抽出を防ぐためには，量子数をある程度大きくしなければならないと考えられる．また，量子数がどのような影響を及ぼすかはデータによってことなることも考えられた．

参考文献

1) Scott C. Bunce, Meltem Izzetoglu, Kurtulus Izzetoglu, Banu Onaral, and Kambiz Pourrezaei. Functional near-infrared spectroscopy. *IEEE ENGINEERING IN MEDICINE AND BIOLOGY MAGAZINE*, Vol. 25, No. 4, pp. 54–62, 2006.

2) Takuma Nishii, Tomoyuki Hiroyasu, Masato Yoshimi, Mitsunori Miki, and Hisatake Yokouchi. Similar subsequence retrieval from two time series data using homology search. *Systems Man and Cybernetics, 2010 IEEE International Conference on*, pp. 1062–1067, 2010.

3) T.F. Smith and M.S. Waterman. Identification of common molecular subsequences. *J. Mol. Biol*, Vol. 147, pp. 195–197, 1981.

4) Abdullah Mueen, Eamonn Keogh, and Qiang Zhu Sydney Cash. Brandonwestover. exact discovery of time-series motifs. *SDM*, pp. 473–484, 2009.

5) Eugene G. Shpaer, Max Robinson, David Yee, James D. Candlin, Robert Mines, and Tim Hunkapiller. Sensitivity and selectivity in protein similarity searches: A comparison of smith waterman in hardware to blast and fasta. *GENOMICS*, Vol. 38, pp. 179–191, 1996.

6) Herbert A. Sturges. The choice of a class interval. *Journal of the American Statistical Association*, Vol. 21, No. 153, pp. 65–66, 1926.

7) 小林瑞季, 小林一郎. 複数の時系列データの比較に基づく言語化の試み. *The 26th Annual Conference of the Japanese Society for Artificial Intelligence*, 2012.

SVMと学習データ選択による病理データのクラス分類アルゴリズムの検討

大堀 裕一
Yuichi OHBORI

背景：研究の背景 癌の診断は病理診断によって行われるが，高度な技術であるため診断補助が求められる．
研究目的：研究の大目的　医師が良性か悪性かの診断を行う際に補助となる情報を提供することである．
発表の位置づけ：小目的　病理データを用いて悪性であるか良性であるか正確な情報を得ることができるか確認する．
方法：提案手法　学習データの組み合わせを選択し，悪性と良性の領域と判断不可能領域の 3 クラスに分離する．
結果：実験結果　乳癌の腫瘍データを用いて実験を行った結果，悪性と良性の領域と判断不可能領域の 3 つに分離できた．

1　はじめに

癌の診断において，患者の腫瘍が悪性であるか良性であるかという判断は病理診断によって行われる．病理診断とは，生体から摘出した組織を薄く切りだして染色し，顕微鏡で観察することによって病変の有無や種類を診断する方法である．病理診断は，病理学の知識や医師の経験によって診断を行う専門性の高い技術である．そのため，診断する医師によって結果が異なるといったことや，医師の経験の違いによって精度が異なるといった問題点が存在する．そこで過去の病理診断によって得られた患者の腫瘍のデータを使用して，医師がある患者の腫瘍が良性か悪性か診断する際に補助となる情報を提示するツールが求められている．しかし，悪性腫瘍と良性腫瘍のデータの分布に重複する領域がある場合，良性と悪性の領域に分けても両方の領域において誤分類が存在し，患者の腫瘍データが入力された際良性，悪性どちらであるか正確に判断できないという問題が存在する．そこで本研究では，悪性，良性という 2 つのクラスに完全に分離することが困難なデータに対して，過去のデータから良性または悪性であると確実に判断できる領域とどちらか判断できない領域に分ける事を提案する．すると，今まで正確に判断できなかったデータに対しても，一部領域においては確実に判断することが可能となる．本研究では，患者の乳房から採取した腫瘍のデータを使用した．腫瘍のデータとは，ある患者の腫瘍が良性または悪性であるかを示し，その腫瘍の病理画像から得られたいくつかの特徴量を持つものである．

本稿では，従来の 2 値分類学習法ではクラスの完全な分離が困難であり，未知データのクラスの正確な判断ができない場合においても，一部領域において判断可能とする提案手法について述べた後，提案手法を用いて判断可能領域と判断不可能領域に分離することが可能であるか実験を行った．その結果，提案した領域に分離できたことを確認した．

2　良性悪性のクラス分類

2.1　病理画像の特徴量を用いた分類

本研究では，人体から摘出した腫瘍の病理画像から得られたデータを使用している．このデータは，ある腫瘍が良性であるか，悪性であるかというクラスと，その腫瘍の病理画像から抽出したいくつかの特徴量を値として持つ．ある病理画像から得られたデータは特徴空間と呼ばれる d 次元空間の 1 点として表現できる．この d 次元ベクトルを特徴ベクトルと呼ぶ．上記のデータを良性と悪性の 2 クラスに分類する問題として，入力と出力間の関数を，与えられたデータから学習する方法を考える．学習とは学習データを利用して，未知の特徴ベクトルがどちらのクラスに属するか判定する関数を求めることである．本研究では，患者の腫瘍病理画像の特徴量を入力とし，その腫瘍が良性であるか悪性であるかということを出力とする関数を学習することを考える．学習は特徴空間上で，2 クラスの識別線を決定することに相当する．

2.2　SVM

本研究では，識別能力に優れている SVM(Support Vector Machine)[1) 2)] をパターン認識手法として用いる．SVM の識別線から最も近いデータをサポートベクトル，識別線からサポートベクトルまでの距離をマージンという．SVM における学習とは，図 1 のようにマージンが最大となるような線形識別線を計算する事に相当する．SVM は汎化能力に優れており，カーネルトリックによって非線形に拡張することが可能である [3) 4) 5)]．

3　提案手法

3.1　提案手法の概要

Fig.2 のような，良性のデータと悪性のデータが重複する部分が存在する学習データの場合，完全なクラス

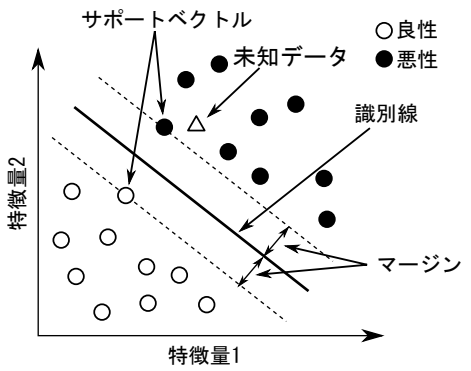


Fig. 1　2 次元データにおける SVM

の分離が困難であり，従来の方法で良性と悪性のクラスに分けたとしても誤った分類が存在してしまい，診断の精度としては信頼性が低下する．そこで，悪性と良性のデータが重複した領域は，クラスの判断が不可能な領域として，悪性と良性の領域とクラスが重複している判断不可能領域の 3 つのクラスに分ける手法を提案する．これによって，一部領域において良性または悪性であるか過去のデータからは正確に判断が可能となる．さらに，悪性と良性の領域をを広くすることが求められる．これは，提案した領域に分離できるように学習データの組み合わせを選択することで実現する．例えば，悪性と判断できる領域と判断不可能な領域に分離したい場合，悪性の学習データの組み合わせを選択する．Fig.3 は悪性のデータを判断可能領域として選択した例である．するとクラスの分離が不可能であったデータが，Fig.3 のように過去のデータから悪性と判断できる領域と，どちらか判断できない領域に分離することができる．同様に，良性と判断できる領域も分離を行う．学習データ選択の方法としてはデータの組み合わせ最適化問題と捉え，最適化アルゴリズムによって行った．最適化アルゴリズムには，組み合わせ最適化問題に適した遺伝的アルゴリズム (Genetic Algorithm:GA)[6) 7)] を用いた．GA は交叉，突然変異といった遺伝的操作を行う最適化アルゴリズムであり，問題に適した解を効率よく探索することができる．

3.2　提案アルゴリズム

以下に提案するアルゴリズムの流れを説明する．

Step 1　[初期個体生成]
悪性の領域を分離する場合は悪性の学習データの組み合わせを，良性の領域を分離する場合は良性の学習データの組み合わせをランダムで多数生成する．

Step 2　[評価]
生成した学習データの組み合わせ一つ一つに対して SVM による分類を行い評価を行う．

Step 3　[選択]
学習データの組み合わせの評価が高いほど選択されるようにする．

Step 4　[交叉，突然変異]
選択された学習データの組み合わせに対して，交叉，突然変異といった遺伝的操作を行い新たな学習データの組み合わせを生成する．

Step 5　[終了条件]
Step2〜Step4 を終了条件に達するまで繰り返す．

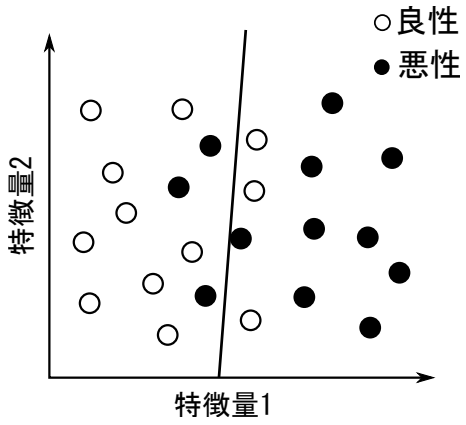


Fig. 2　学習データの例

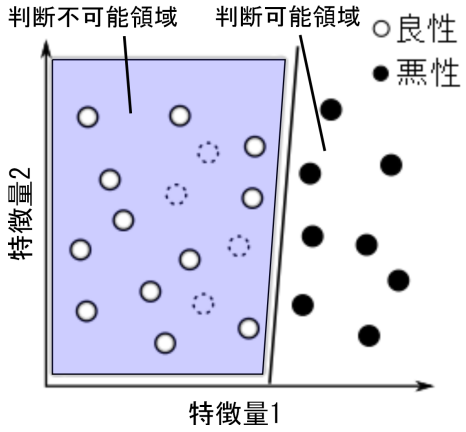


Fig. 3　提案手法の概要

3.3　最適化問題の定式化

n 個のデータ $(\boldsymbol{x_i}, y_i)$,$i = 1...n$ が与えられているとする．このとき，$x_i \in R^n$ は特徴ベクトルであり，$y_i \in (1, -1)$ はそれぞれの特徴ベクトルに対応するクラスである．設計変数を学習データの組み合わせとして，選択するデータを 1，選択しないデータを 0 でデータ長の 2 値ビット配列で表し，Fig.4 のように学習データに対応させる．n 個の中から k 個が選択されたとき，学習データ数は k 個となる．そこで最適化問題としては選択された学習データに対して誤識別数が 0 という制約条件で，選択される学習データ数 k を最大化する問題となる．誤識

別の度合いとしては，識別線から誤識別データまでの距離の総和として表す．識別線から誤識別データまでの距離は SVM Confidence Margin によって算出される．誤識別を以下のように定義する．

$$m(y, f(\boldsymbol{x})) = \begin{cases} 1 & \text{if } yf(\boldsymbol{x}) < 0 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

学習データ数が k 個の場合，学習データに対する誤識別率 Err は式 (3) のように表される．

$$Err = \frac{1}{k}\sum_{i=1}^{k} m(y_i, f(\boldsymbol{x}_i)) \tag{2}$$

また識別線から誤識別データまでの距離の総和は以下のように表される．

$$D = \sum_{i=1}^{k} |f(\boldsymbol{x}_i)| m(y_i, f(\boldsymbol{x}_i)) \tag{3}$$

すると目的関数，制約条件は以下の式で表される．制約条件を満たさない場合はペナルティとして減じ，誤識別率が 0 となるようにする．

$$\text{maximize } O = k - \sum_{i=1}^{k} |f(\boldsymbol{x}_i)| m(y_i, f(\boldsymbol{x}_i)) \tag{4}$$

$$\text{subject to } Err = 0 \tag{5}$$

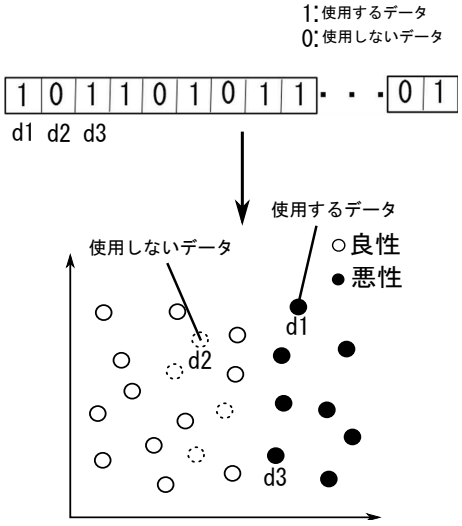


Fig. 4　データ選択の例

4　実験

4.1　実験概要

今回の実験の目的は，クラスの完全な分離が不可能なデータに対して提案手法を用いて悪性，良性と判断できる領域と，どちらか判断不可能な領域の 3 クラスに分けられるかを確認することである．まず，2 次元データにおいて従来の SVM と提案手法を行い結果を比較する．次に，使用する特徴量を 3 つとして 3 次元データに対しても提案手法の適用が可能であるか確認を行い，2 次元の場合と比較した．GA で用いたパラメータを Table.2 に示す．また，SVM による識別線は 2 次元多項式とした．

Table. 1　GA のパラメータ

パラメータ	値
世代数	500
個体数	500
選択手法	トーナメント選択
トーナメントサイズ	4
交叉方法	一様交叉
交叉率	0.9
突然変異率	0.01
エリート数	1

4.2　実験データ

データセットは，UCI Machine Learning Repository の breast cancer[8] のデータを用いた．このデータは乳房組織から腫瘍を採取し，その腫瘍が良性であるか悪性であるかというクラスを示し，腫瘍の病理画像から抽出した 10 個の特徴量を有している[9]．569 個のデータの中から 455 個を対象データとした．このうち良性は 269 個，悪性は 186 個である．Fig.5 に 2 次元データの分布を示す．breast cancer データの特徴量を Table.1 に示す．

Table. 2　腫瘍データの特徴量

	特徴量
1	radius
2	texture
3	perimeter
4	area
5	smoothness
6	compactness
7	concavity
8	concave points
9	symmetry
10	fractal dimension

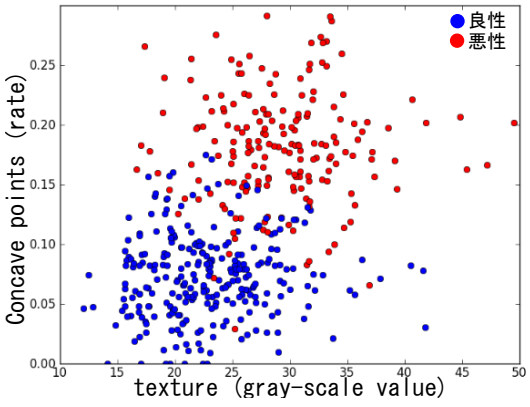


Fig. 5　腫瘍データ

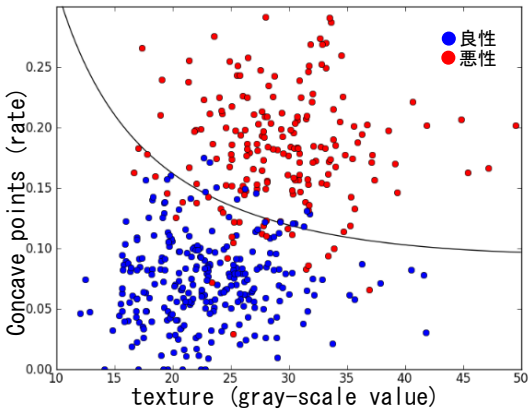


Fig. 6　従来の手法の結果

4.3　実験結果

4.3.1　2 次元データによる実験結果

2 次元データの特徴量は Texture と Concave Points とした．Fig.6 に従来の SVM で実験を行った結果，Fig.7 に提案手法を用いた実験結果を示す．従来の方法で良性のクラスと悪性のクラスに分けたとき，良性の領域において誤識別数が 13 個，悪性領域において誤識別数が 29 個となった．提案手法を用いた場合，悪性，良性の領域において誤識別数は 0 となった．また削除された悪性データ数は 57 個，削除された良性データ数は 130 個となった．

Table. 3　2 次元データにおける従来手法の結果

良性領域の誤識別数 (個)	13
悪性領域の誤識別数 (個)	29

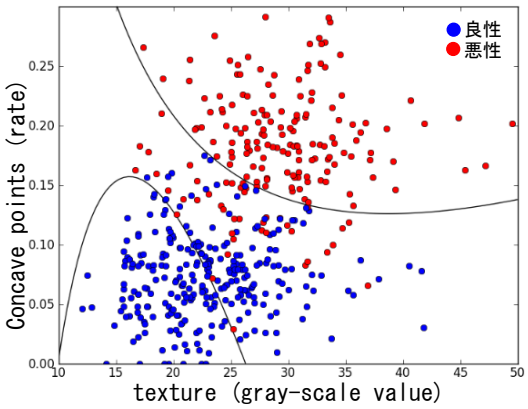


Fig. 7　提案手法の結果

Table. 4　2 次元データにおける提案手法の結果

悪性領域の誤識別数 (個)	0
削除された悪性データ (個)	57
良性領域の誤識別数 (個)	0
削除された良性データ (個)	130

4.3.2　3 次元データによる実験結果

3 次元データにおいて特徴量は Texture, Concave Points, Radius とした．従来の SVM で良性と悪性のクラスに分けた場合，良性の領域において誤識別数が 6 個，悪性の領域において誤識別数が 45 個となった．提案手法を用いた場合，悪性の領域において誤識別数は 0 個となった．また削除された悪性データ数は 31 個となった．

Table. 5　3 次元データにおける従来手法の結果

良性領域の誤識別数 (個)	6
悪性領域の誤識別数 (個)	45

Table. 6　3 次元データにおける提案手法の結果

悪性領域の誤識別数 (個)	0
削除された悪性データ数 (個)	31

5　考察

従来の方法で良性と悪性の領域に分けた場合，良性と悪性の領域両方において誤った分類が存在するため，新たに診断する患者の腫瘍データが入力されるとき，その

腫瘍が良性であるか悪性であるか正確な判断が困難となる．提案手法を用いた場合は，悪性と良性の領域において誤識別が 0 となった．従って，患者の腫瘍のデータがこのどちらかの領域に入力された場合，過去のデータからは悪性または良性であると判断できる．以上から，悪性，良性と判断できる領域とどちらであるか判断不可能な領域の 3 クラスに分離することができたと考えられる．また 2 次元データにおいての提案手法の結果と 3 次元データにおいての提案手法の結果を比較すると，削除された悪性データ数は 3 次元の方が小さい結果となった．このことから 3 次元データで提案手法を行った方が判断可能領域は広くなると考えられる．

6 まとめ

本稿では，医師が患者の腫瘍の診断を行う際の補助となる情報を提示するため，患者の腫瘍の病理画像から得られたデータを学習して，新たに診断を行う患者の腫瘍が良性であるか，悪性であるか判断する方法を提案した．しかし，良性と悪性の 2 つのクラスに完全に分離することが困難な場合，SVM でこの 2 つのクラスに分けたとしても信頼性が低下する．そこでこういったデータに対して，どちらのクラスであるか過去のデータから判断が可能な領域と，判断不可能な領域に分ける手法を提案した．この問題はデータの組み合わせ最適化問題と捉え，学習データの組み合わせを設計変数とし，誤識別が 0 という制約条件で学習データ数を最大化する最適化問題とすることで実現した．breast cancer の病理画像のデータを用いて実験を行った結果，悪性と良性の領域と判断不可能領域の 3 クラスに分けることができた．また 2 次元データで行った場合と 3 次元データで行った場合を比較すると，後者の方が判断可能領域が広くなることが示唆された．

参考文献

1) Vladimir Vapnik, ”The support vector method of function estimation”, in J.A.K. Suykens and J.Vandewalle Nonlinear Modeling, Advanced Black-Box Techniques, Kluwer Academic Publishers, Boston, pp.55-85, 1998.

2) Vladimir Vapnik, ”Statistical learning theory”, John Wiley, New York, 1998.

3) Terrence Furey, Nello Cristianini, Nigel Duffy, David Bednarski, Michel Schummer and David Haussler, ”Support vector machine classification and validation of cancer tissue samples using microarray expression data”, Bioinformatics, vol.16, no.10, pp.906-914, 2000.

4) Sujun Hua and Zhirong Sun, ”Support vector machine approach for protein subcellular localization prediction”, Bioinformatics, vol.17, no.8, pp.721-728, 2001.

5) Nello Cristianini, John Taylor, ”An Introduction to Support Vector Machines:And Other Kernel-Based Learning Methods”, Cambridge University Press, 2000.

6) David Goldberg, ”Genetic algorithms in search, optimization and machine learning”, Addison Wesley, 1989.

7) Darrel Whitley, ”A genetic algorithm tutorial”, Statistics and computing, vol.4, no.2, pp.65-85, 1994.

8) UCI Machine Learning Repository. [Online]. Available: `http://archive.ics.uci.edu/ml/datasets.html`

9) Nick Street, William Wolberg, Olvi Mangasarian, ”Nuclear Feature Extraction For Breast Tumor Diagnosis”, International Symposium on Electronic Imaging, Science and Technology, vol.1905, pp.861-870, 1993.

角膜内皮細胞のモデル画像におけるパラメータ制御手法

上堀 聖史
Kiyofumi UEHORI

背景：専門家による角膜内皮細胞の正常性判断の基準において，定量的かつ一般的な基準が定められていない．
研究目的：角膜内皮細胞の正常性判断の客観的基準を定め，より確実な角膜内皮細胞移植を行う再生医療への支援．
発表の位置づけ：パラメータが制御可能な細胞のモデル画像を生成し，専門家に画像提示を効率的に行える仕組みを構築する．
方法：GA または最急降下法を用いた細胞のモデル画像のパラメータ制御を可能にするための手法．
結果：両手法の計測結果では，GA はより高速な画像生成が行え，最急降下法はより精度の高い画像生成が行えた．

1　はじめに

近年，病気や機能不全に陥った組織や臓器を再生する再生医療が発展しており，角膜内皮における再生医療もそのうちの一つである [1]．角膜の 5 層構造の最も内側である内皮層に存在する角膜内皮細胞は，角膜の透明性を維持するのに重要な役割を果たしている．角膜内皮での再生医療では，ドナー角膜または患者の角膜内皮から細胞を *in vitro* で培養し，それを再び患者の角膜内皮に移植することで，角膜の機能回復を目指している．この際，移植する細胞の状態が，移植後の角膜の機能を左右する．

角膜内皮細胞は，状態により細胞密度や形態，大きさなどの特徴が異なるものとなる．実際に専門家は，角膜内皮細胞が良好な状態かどうかの判断において，細胞密度，細胞の大きさの偏差，全細胞数における六角形細胞数が占める割合を判断指標としている．

しかし，専門家は細胞の正常性を主観的判断のみで診断しており，すなわち専門家自身の知識と経験のみしか判断材料がない．細胞の正常性を定量的に示し，それを複数の専門家間で一般化することで，細胞の正常性判断基準を確立する．専門家がその基準を考慮した診断を行うことで，細胞の正常性判断がより確実なものになり，角膜内皮細胞移植の効率の向上化に繋がる．

そこで本研究は，細胞の正常性の客観的な基準抽出を行う方法を構築し，より確実な角膜内皮細胞移植の支援となることを目指している．そのために，専門家にパラメータが異なる細胞画像を提示していき，判断を基に，どの程度のパラメータであれば正常性であるかを抽出するといった方法を検討している．この方法においては，実際の細胞画像を模擬したモデル画像を採用する．このモデル画像は，細胞の中心点を与えることで生成するが，この中心点の位置を制御することで，モデル画像のパラメータを制御できる．しかし，中心点の位置を制御するためにはかなりの計算を要するため，目的とする中心点位置を少ない計算回数で求めるための手法が必要である．

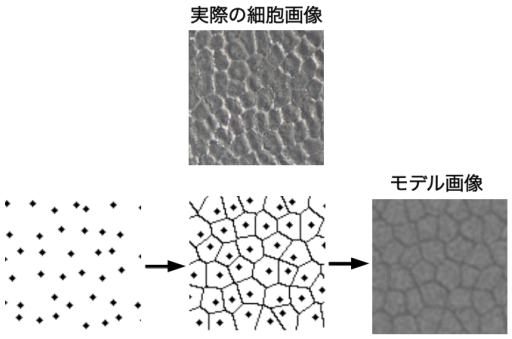


Fig. 1　モデル画像の生成過程

本稿では，モデル画像のパラメータを制御するために 2 つの最適化手法を適用し，その結果について述べる．

2　本手法の要件

2.1　提示する細胞画像

本手法においては，定めた特徴量をもつ細胞画像を提示する仕組みが必要である．この際，実際の細胞画像を用いる場合を考えると，ある特徴量をもつ細胞画像を提示し，次に提示するためのその近傍値をとる細胞画像が存在しないことも想定される．よって，実際の細胞画像を本手法に適用するのは適切でないと考えられる．これを解決するために，実際の細胞画像を模擬したモデル画像を採用する．モデル画像は，Fig. 1 のように，細胞の中心点を与え，その中心点を基にボロノイ分割 [2] することで画像を領域分割する．その後，実際の細胞画像と比較的同等なグレースケール値を割り当てることで，細胞を模擬する．この構築方法では，細胞の中心点が細胞数に相当し，その中心点の位置関係により，細胞の形態を変化させることができる．ゆえに，この中心点の数や位置を制御することで，特定の特徴量をもつ細胞画像を構築することが可能である．

2.2　定めた特徴量の細胞画像生成

前節において提示する細胞画像を述べたが，さらに定めた特徴量をもつ細胞画像を生成し提示する手法を考える．留意する点として，生成する細胞画像の特徴量は，生成する際の細胞の中心点の数や配置に依存している．特徴量から細胞の中心点の位置を逆算するための手法を考える必要がある．ゆえに，定めた特徴量をもつ細胞画像の生成を，細胞の中心点配置の最適化問題と捉えることができる．また，専門家である人を対象とする方法であるため，ある程度迅速な画像提示が求められる．しかし，細胞の中心点配置の組み合わせは膨大な数があり，それなりの計算時間を有する．本稿では，高速化を図るため，最適化手法である遺伝的アルゴリズム (Genetic Algorithm：GA)，および勾配法の一つである最急降下法を本問題に適用した．

2.2.1　GA

GA は，生物の進化過程を工学的にモデル化した学習的アルゴリズムである [3]．自然界における生物の進化過程では，ある世代を形成している個体の集合の中で環境に適応した個体が高い確率で生き残り，次の世代に子を残す．このメカニズムをモデル化し，環境に対して最も適応した個体，すなわち目的関数に対して最適値を与えるような解を探索する．また，細胞の中心点の位置は，画像における平面空間上の座標によって表現される．よって，今回は個体をビットストリングではなく実数値で表現する実数値 GA を用いる．

2.2.2　最急降下法

最急降下法は，関数の傾きのみから関数の最小値を探索する勾配法のアルゴリズムの一つである．最急降下法のアルゴリズムを以下に示す [4]．ここで，x を n 次の変数，f を目的関数とする．

最急降下法のアルゴリズム

1. 探索の出発点 $x(0)$ を定め，$k=0$ とおく．
2. $\nabla f(x_1(k), \cdots, x_n(k))=0$ ならば終了．
   違う場合，
   $d(k)=-\nabla f(x_1(k), \cdots, x_n(k))$ とおいてステップ 3 へ．ただし，$\nabla f=[\frac{\partial f}{\partial x_1}, \cdots, \frac{\partial f}{\partial x_n}]$.
3. ステップ幅 $\alpha(k)$ を求め，次の探索点
   $x(k+1)=x(k)+\alpha(k)d(k)$ を定める．
   $k=k+1$ とおいて，ステップ 2 へ．

3　手法の評価実験

3.1　実験目的

本稿で取り上げた手法である GA と最急降下法を用い，評価実験を行った．実験の目的としては，それぞれの手法により，目標とする特徴量をもつ細胞画像を生成するのに要する時間を確認することである．今回は，細胞の正常性判断に関係する，細胞面積の標準偏差 (Standard Deviation：SD) を対象に行った．なお，細胞面積 SD は，ラベリング処理[1]により算出する．

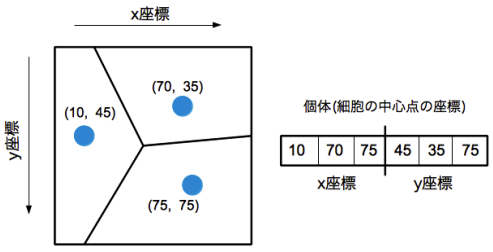


Fig. 2　個体表現

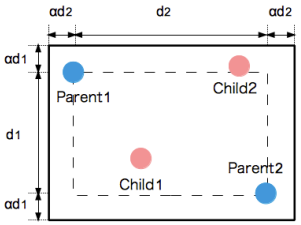


Fig. 3　BLX-α

3.2　実験パラメータ

Fig. 2 の左のように細胞の中心点が配置された場合，右のように個体が表される．

GA および細胞画像のパラメータを Table. 1，Table. 2 に示す．また，両手法において初期値はランダムとした．本問題でも個体は，細胞の中心点の x 座標，y 座標である．よって，個体の表現は Fig. 2 のようになる．交叉 BLX-α は，Fig. 3 のように，二つの親個体からそれぞれ子個体の発生範囲を決めて子個体を生成する手法である．親個体実数ベクトルの区間 di を両側に α di だけ拡張した区間からランダムに子個体を生成する．終了条件は，目的関数である，入力した特徴量と生成した画像に対して算出した特徴量の絶対値の差が 0 になったときとする．なお，最急降下法における目的関数は，絶対値の差を二乗したものとする．

Table. 1　画像パラメータ

パラメータ	値
画像サイズ	100 × 100 ピクセル
細胞の中心点数	50

[1] 二値化処理された画像において，白 (黒) が連続した画素に同じ番号を割り振る処理

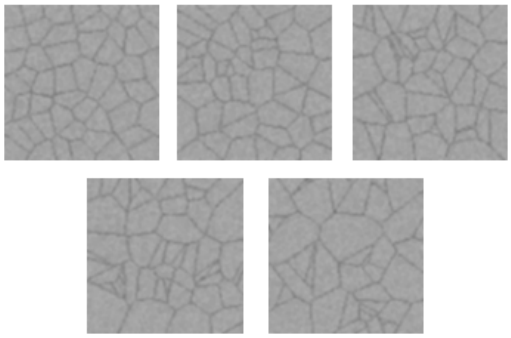


Fig. 4　生成した画像生成例 (上段左から細胞面積 SD：40，80，120，下段左から細胞面積 SD：160，200)

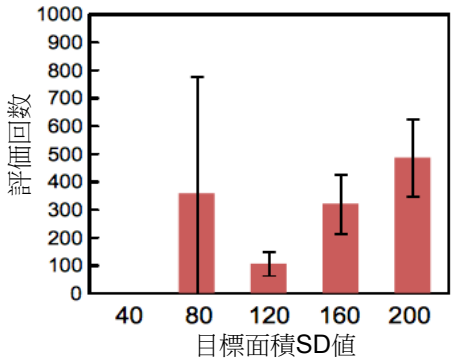


Fig. 5　GA による画像生成における評価回数

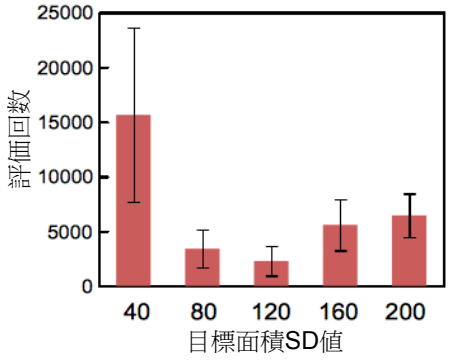


Fig. 6　最急降下法による画像生成における評価回数

3.3　実験結果

今回では，細胞面積 SD 値が 40，80，120，160，200 を目標値として定め，それぞれの目標値の細胞画像を生成することを試みた．また，それぞれの目標値に対して，10 試行ずつ行った．各目標値で生成された細胞画像例を Fig. 4 に示す．

GA で画像を生成した結果を Fig. 5 に示す．一世代の実行時間はおよそ 0.7 秒であった．目標値が 40 のときは，10 試行中一度も収束に至らなかった．

最急降下法で画像を生成した結果を Fig. 6 に示す．最急降下法では，GA と比較すると収束速度は遅かったが，各目標値において 10 試行全て収束した．

3.4　考察

3.4.1　収束速度と初期値の関係

Fig. 5 において，目標値が 120 の際に最も収束が早く，次いで 80, 160, 200 という結果となっている．これは最急降下法での結果である Fig. 6 も同様の結果となっている．これらの理由として，初期値が関係すると考えられる．今回，両手法においては，細胞の初期の中心点はランダムに配置したが，ある程度，細胞面積 SD の初期値は限られた範囲に生成されているのではないかと考えた．そこで，細胞の中心点をランダムに設定した場合の細胞面積 SD の算出値を確認した．細胞面積 SD 値を 100 回算出した結果を Table. 3 に示す．平均値が 107.49，標準偏差が± 17.51 となった．この結果により，算出値の平均値と近い値である目標値において高速な収束を示すことが考えられる．Fig. 5，Fig. 6 においても，目標値 120 の際に最も高速な収束を示している．

Table. 2　GA の各種パラメータ

パラメータ	値
個体数	50
世代数	600
交叉率	1.0
突然変異	なし
遺伝子長	100
交叉方法	BLX-α
α の値	0.5
選択方法	トーナメント選択
トーナメントサイズ	2

3.4.2　両手法における中心点の更新方法の相違

Fig. 5 より，GA において目標値 40 において収束しなかった．目標値 40 は，Table. 3 の結果を踏まえると，初期値より小さい値になることが考えられる．これにより，目標値が初期値より小さい場合に収束が困難になることが考えられる．このことは，細胞の中心点の更新方法が関係すると思われる．目標値 40 に収束しなかった GA での中心点の更新方法は，交叉である BLX-α によるものである．BLX- α は，二つの親個体からそれぞれ子個

Table. 3　100 試行における細胞面積 SD の算出値

最大値	最小値	平均値 (mean ± SD)
161	73	107.49 ± 17.51

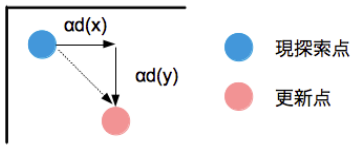


Fig. 7　探索ベクトルによる細胞の中心点の更新

体の発生範囲を決めて子個体を生成する手法である．この子個体は，発生範囲を決めて一様乱数により決定するという，ランダムな要素を含む．ゆえに，細胞面積 SD を初期値より拡大していくことは可能であるが，縮小させていくことが困難だったのではないかと考えられる．Fig. 5 の目標値 80 において，標準偏差のばらつきが他の目標値と比較し，かなり大きいことがわかる．これは初期値が 80，または 80 以下に生成された場合は容易に収束するが，初期値が 80 より大きく生成された場合は評価回数が多くなることが考えられる．Fig. 4 の目標値 40 の画像例のように，細胞面積 SD が小さい場合，細胞の中心点はある程度規則的に並べられる状態になる．この細胞の中心点の規則的な配置を探索するのは，ランダムな要素を含む場合では困難だったのではないかと考えられる．一方，最急降下法では，Fig. 7 に示すように，ベクトル方向に細胞の中心点を更新する．この方法では直接的に探索するため，ランダムな要素は含まれておらず，収束する確率が高かった．以上により，細胞の中心点の更新方法は，ベクトル方向で探索する方が適切ではないかと考えられる．

3.4.3　最急降下法のステップ幅

最急降下法の収束速度の高速化について述べる．これは，最急降下法におけるステップ幅 α が関係していると思われる．今回実装した最急降下法では，ステップ幅 α を固定値としていた．しかし，ステップ幅 α のとる値によっては，目的関数値が増加することもあるため，評価毎に定める必要がある．実際，目的関数 $f(x)$，現探索点 x_k，探索ベクトル d_k，ステップ幅 α とすると，関数

$$\phi\ (\alpha) = f(x_k + \alpha\ d_k),\quad \alpha > 0 \tag{1}$$

を考え，一変数の最小化問題

$$min\{\ \phi\ (\alpha)|\ \alpha > 0\} \tag{2}$$

の解をステップ幅 α に定める方法がある [5]．本実験では 10 試行においては収束が見られたが，ステップ幅 α が固定値であると収束せず発散する可能性も考えられる．よって，ステップ幅を動的に定めることで，さらに高速かつ精度の高い探索が可能であることが期待される．

4　課題点

今後の課題として，まず高速かつ精度の高い手法を確立することである．今回の実験結果では，収束の高速性に関しては GA の方が良く，精度としては最急降下法の方が高かった．それらの要因を調査し，適切な手法を確立することが一つ目の課題点である．さらに，本稿では細胞面積 SD のみを考慮したものであった．今後は，さらにそれ細胞密度や六角形細胞数の割合といった指標も考慮した，多目的な画像生成手法を構築することが必要である．

5　まとめ

本研究では，角膜内皮における正常性判断の客観的基準を抽出手法を構築し，より確実な再生医療を目指すことを目的としている．この際，専門家に特徴量の異なった細胞画像提示を行い，正常性といえる特徴量基準を抽出する方法をとる．本稿では，専門家に細胞画像を効率的に提示できる仕組みについて検討した内容を述べた．今回は，GA および最急降下法を用いた画像生成手法について検討し，それらの細胞画像生成における評価実験を行った．実験結果として，両手法を比較したところ，GA ではより高速な画像生成が行え，最急降下法では，より精度の高い画像生成が行えることが考えられる．今後，これらの結果を踏まえ，効率的な手法を確立し，さらに本稿で示した細胞面積 SD のみの単目的ではなく，多目的に行えることが課題である．

参考文献

1) Kazuo TSUBOTA and Shin HATO. Corneal disease and regenerative medicine. *TRENDS IN THE SCIENCES*, Vol. 15, No. 7, pp. 8–13, 2010.

2) 神田毅, 杉原厚吉. 最小 2 乗法によるボロノイ図あてはめ. 数理解析研究議事録, Vol. 1297, pp. 125–135, 2002.

3) 樋口哲也, 北野宏明. 遺伝的アルゴリズムとその応用. 情報処理学会, Vol. 34, No. 7, pp. 871–883, 1993.

4) Doron BLATT Alfred HERO and Hillel GAUCHMAN. A convergent incremental gradient method with a constant step size. *SIAM Journal on Optimization*, Vol. 18, No. 1, pp. 29–51, 2007.

5) Ya xiang YUAN. A new stepsize for the steepest descent method. *Journal of Computational Mathematics*, Vol. 24, No. 2, pp. 149–56, 2006.

角膜内皮細胞画像特徴量抽出システムにおける学習画像作成の検討

布川 将来人
Sakito NUNOKAWA

背景：角膜再生医療において，細胞の培養状態を確認する際に画像処理の知識が必要であるなどの負担が存在する．
研究目的：角膜内皮細胞画像から特徴量を自動で抽出する事によって，細胞状態の確認を支援する．
発表の位置づけ：角膜内皮細胞画像の特徴量自動抽出システムにおいて用いる学習画像作成の簡易化の検討
方法：細胞画像の各細胞領域に対して点をプロットし前処理及び watershed 法を用いる．
結果：seeded-watershed によって作成した画像が学習画像として有効であることが確認できた．

1　はじめに

昨今の高齢化社会における問題のひとつとして角膜障害があげられる．この角膜障害の中でも，角膜内皮に起因する障害は治療が困難であり，現状ではドナーが提供した角膜を移植する事でしか治療を行うことができず，ドナー不足や拒絶反応などが問題とされている．この角膜障害に対する新たな治療法として，角膜内皮再生医療が現在研究されている [1]．

角膜内皮再生医療では，患者やドナーから得た角膜内皮細胞を体外において培養する．そして培養によって得られた角膜内皮を患者に移植することで治療を行う．この治療方法によって，一人のドナーから得た角膜内皮を複数の患者に移植することができ，また患者から得た角膜内皮細胞を培養，移植する事ができれば拒絶反応などの問題が解消されるのではないかと期待されている．しかし，角膜内皮再生医療においても様々な問題点が存在する．その中の一つとして細胞の培養状態の確認があげられる．再生医療では，細胞を培養して得られた組織を使用する為，それらの細胞がうまく培養できているかを判断する必要がある．角膜内皮再生医療においては，顕微鏡によって撮像した細胞の画像から細胞密度や細胞の形状（正六角形に近いかどうか）などを細胞の特徴量として計測し，それらの特徴量の値を基に培養状態の判断を行っている．これらの特徴量の計測は ImageJ や CellProfiler などの既存の画像処理ソフトを用いて行われているが，これらのソフトを用いるには，画像に適した画像処理の方法（組み合わせ）や，閾値などのパラメータを設定する必要がある [2) 3)]．そのため，これらの設定には画像処理の専門的知識が必要であり，再生医療の研究者や医師の負担となっている．

そこで本研究グループでは，画像処理の知識を必要とせず，角膜内皮細胞画像から細胞密度などの特徴量を自動で計測できるシステムの提案を行っている [4]．

しかし，本システムでは GP を用いる際に学習画像として事前に目的の処理を行った画像の例（目標画像）をユーザーが任意の方法で用意する必要がある．そこで，本稿では学習画像となる目標画像作成のかかる負担を軽減する為に，目標画像作成手法の提案及び検討を行う．本稿では，角膜内皮細胞画像特徴量自動抽出システムの概要，学習データ作成手法の提案，提案手法の評価実験について示す．

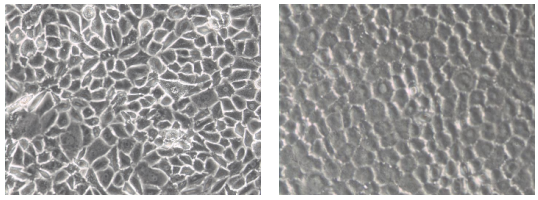


Fig. 1　培養角膜内皮細胞画像

2　角膜内皮細胞画像特徴量自動抽出システム

2.1　システム概要

本研究では角膜内皮細胞画像から，細胞の特徴量を自動で抽出する事を目的としている．しかし，培養された角膜内皮細胞画像は Fig. 1 に示すような画像となっており，直接特徴量（細胞密度，形状）を計測する事が困難である．そこで，本システムでは以下の手順によって特徴量の抽出を行う．

step.1 細胞画像を細胞領域と境界に分割
step.2 細胞領域から特徴量を抽出

2.2　細胞領域分割

一般的に複雑な画像処理も既知の単純な画像処理フィルタを複数組み合わせることで表現することができると考えられる．そこで本システムでは，画像処理の組み合わせを木構造の組み合わせ最適化問題と捉え，木構造の組合せ最適化手法である遺伝的プログラミング（GP）[5] を用いる．

本手法ではフィルタの組み合わせを Fig. 2 に示すような木構造状フィルタによって表現する．Fig. 2 の木構造状フィルタにおいて各ノード（枝）が既存の画像処理

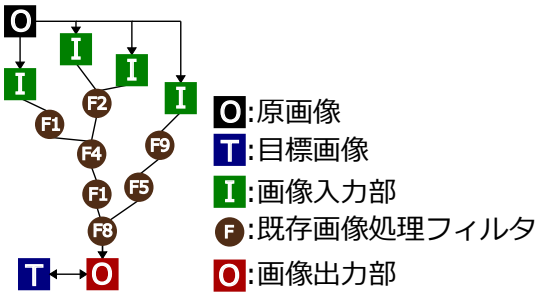


Fig. 2　木構造状フィルタ

フィルタ，各最端ノード（葉）が画像の入力部分，ルート（根）が各ノードによって処理され最終的に画像が出力される部分となる．また，本システムでは GP で用いる学習データとして原画像と目標画像を用いる．原画像とは処理の対象となる画像の一例もしくは一部分であり，目標画像は原画像に対して目的の画像処理を行った結果の例である．また，目標画像はユーザーが任意の方法で作成する必要がある．

本システムでは，自動プログラミング手法である GP を用いることによって，出力画像が目標画像と最も近似するようなフィルタの組み合わせを自動で構築する事ができる．本システムで用いる GP では Fig. 2 のようなフィルタの組み合わせをひとつの個体とし，画像入力部に原画像を入力した際に得られる出力画像と目標画像の一致度を適合度として用いる．一致度は出力画像 $O(x;y)$ と目標画像 $T(x;y)$ の各ピクセル値の差分であり，最大値を 1.0 とする．一致度の計算式を式 (1) に示す．

$$fitness = \frac{\sum_{x=1}^{W_x}\sum_{y=1}^{W_y}|O(x,y)-T(x,y)|}{W_x * W_y * V_{max}} \quad (1)$$

式 (1) において W_x は画像の横サイズ，W_y は画像の縦サイズ，V_{max} は最大階調値（255）を示す．

2.3　特徴量抽出

角膜内皮細胞画像において，細胞サイズ，細胞密度，細胞の形状などの特徴量が重要になってくる．また，細胞密度が高く，形状が正六角形に近い程良好な細胞と判断される．本システムでは step.1 で分割された各細胞領域からラベリング処理を用いて細胞数．細胞サイズ，六角形細胞数を計測する．

3　学習データ作成手法

3.1　概要

2 章で示したシステムでは，Fig. 3 のような学習データセット（原画像・目標画像）を用いることで画像処理を行うが，この目標画像は原画像からユーザーが既存の画像処理ソフトや手作業によって任意に作成する必要がある．学習に用いる原画像は対象画像の一部分である為，画像全体を処理するよりも遥かに少ない労力で行う事ができると考えられるが，よりユーザー負担を少なく目標

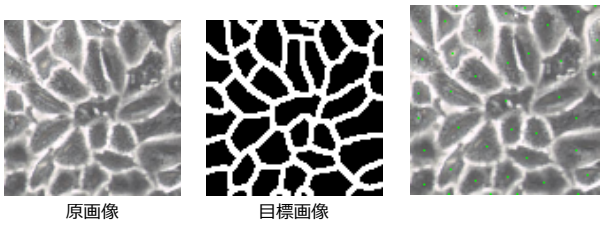


Fig. 3　学習データセット

Fig. 4　各細胞をプロットした画像

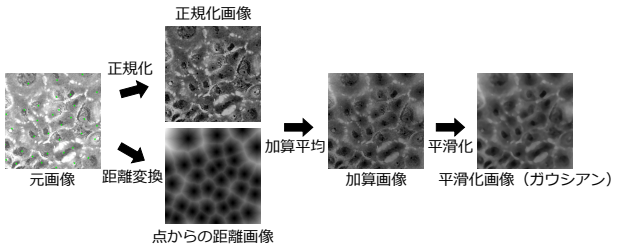


Fig. 5　提案手法における前処理

画像を作成する方法が必要であると考えられる．そこで，本稿ではユーザーが Fig. 4 に示す画像のように，画像の細胞領域に点をプロットする事だけで目標画像を作成できるような手法の提案及び検討を行う．本手法を用いることで，本システムにおけるユーザーの負担を軽減する事ができる．

目標画像の作成方法として以下の 2 種類の方法を提案する．

1. watershed 及び統合法を用いた手法
2. seeded-watershed を用いた手法

watershed とは画像処理でよく用いられる領域分割手法のひとつである [6]．また，ノイズ除去及び原画像にプロットされた点の情報を画像に付加する為に，以下の前処理を行う．

step.1　点からの距離画像を作成
step.2　原画像及び距離画像の画素値を 0〜255 に正規化
step.3　距離画像と正規化画像を加算平均
step.4　加算平均画像に対して平滑化（ガウシアン）

上記の前処理の例を Fig. 5 に示す．

3.2　watershed 及び統合法を用いた領域分割

画像処理における一般的な領域分割の手法として watershed 法があげられる．watershed 法とは seed と呼ばれる領域の開始点を複数設定し，各 seed から領域を拡張していく事によって領域の分割を行う手法である．しかし，一般的に watershed 法では seed 数を自動で決定する為，過分割が多く発生するという問題点がある．

そこで，特定の条件に沿って領域の統合を行う統合法を用いる．統合法とは統合条件を設定し，その統合条件を満たす領域がなくなるまで統合を繰り返す手法である．本手法では，前処理を行った画像に対して watershed+統

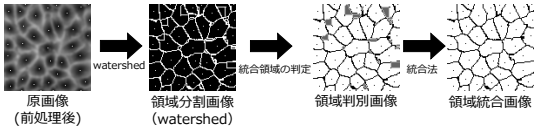


Fig. 6　領域分割+統合法の流れ

合法を用いる．watershed で分割された領域のうち，プロットされた点を含まない領域を統合対象領域とし，統合対象領域が無くなるまで統合を繰り返す．統合先の領域としては統合対象となった領域と隣接する領域の中心点を計算し，中心点の最も近い領域を統合先領域とする．また，統合先領域が点を含まない場合でも統合を行う．watershed 及び統合法の流れを Fig. 6 に示す．

3.3　seeded-watershed を用いた領域分割

seeded-watershed とは 3.2 節で述べた watershed 法の seed 数及び seed 位置を事前に設定する手法である．この seeded-watershed を用いることで，一般的な watershed で発生するような過分割を起こすことなく領域を分割する事が可能である．本手法ではプロットされた点を seed とし閾値を徐々に上げて行きながら，seed もしくは領域に隣接しているピクセルの画素値が閾値以下であればその領域に属するとみなす．閾値にはグレースケール画像の画素値である 0〜255 までの範囲の値を用いる．閾値の上昇幅は 1 とする．また，同じ閾値において複数の領域と隣接し，かつそのピクセルの画素値が閾値以下だった場合そのピクセルを境界とする．

4　評価実験

4.1　実験概要

本実験では，3 章で提案した目標画像作成方法のうちどちらが高精度に目標画像を作成できるか，また提案した方法によって作成した目標画像が GP の学習データとして利用できるかどうかについての検討を行う．

検討方法として，まず提案手法を用いて目標画像を作成し，既存画像処理フィルタ及びドローソフトを用いて任意に作成した目標画像との比較を行う．この段階では画像を定量的に評価するのは困難である為，比較は目視によって行う．

次に，作成した目標画像を学習データとし，GP によってフィルタの構築を行う．構築したフィルタを 2 枚の未学習画像に適応し，得られた領域分割画像の細胞数を比較する事によって提案手法の有効性を確認する．

4 枚の細胞画像（346 × 260[pixel]）を対象の画像とし，各対象画像から 100 × 100[pixel] で切り出した画像を学習データの原画像，原画像とは異なる部分を 100 × 100[pixel] で切り出した画像を未学習画像として用いる．また提案手法で用いる原画像は各細胞領域に緑の点をプロットしたものを用いる．対象画像，点をプロットした

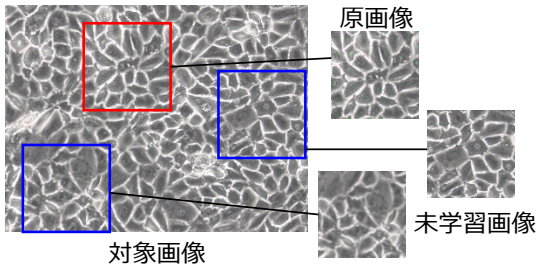


Fig. 7　実験使用画像（一例）

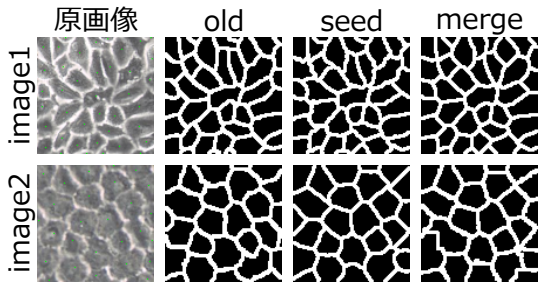


Fig. 8　提案手法適用結果

原画像及び未学習画像の一例を Fig. 7 に示す．対象とした 4 枚の細胞画像はカニクイザルの角膜内皮を培養し，顕微鏡の倍率 200 倍で撮影された培養角膜内皮細胞画像である．GP のパラメータ設定及び使用フィルタは，システムの提案及び検討を行った際に使用したものと同様の設定で行った [4]．

4.2　実験結果

4.2.1　提案手法を用いた目標画像の作成

「watershed+統合法」と「seeded-watershed」の両手法で作成した目標画像の一例を Fig. 8 に示す．Fig. 8 及び以降の図表において，old は任意に作成した画像であり，seed は seeded-watershed，merge は watershed+統合法によって得られた画像である．目視による比較で，seed，merge 共に任意に作成した目標画像に近い画像を生成出来ている事が確認できた．

merge による作成画像において細胞領域の一部分が崩れている部分が存在するが，これは統合過程において統合先の領域が適切な領域で無かった為だと考えられる．現状，統合先の領域は各領域の中心点の距離を用いているが，領域の画素値などを用いる事でより正確な統合が行える可能性が考えられる．また，提案手法で作成した画像は任意に作成した画像に比べ，領域の細かな特徴を反映できていない部分が存在する．これは，前処理で行った距離画像の付加及び平滑化による影響によって原画像の細かな情報が除去された為だと考えられる．しかし，これらの前処理を行わなければ適切な分割（統合）を行う事は困難であるため，前処理の度合についても検討する必要があると考えられる．

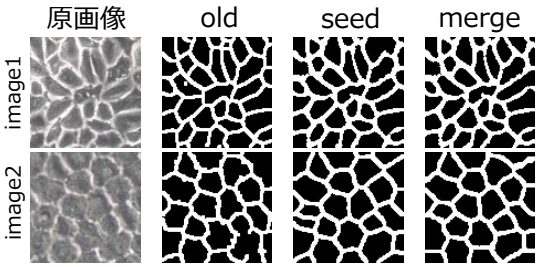


Fig. 9　GP 適用結果（学習画像）

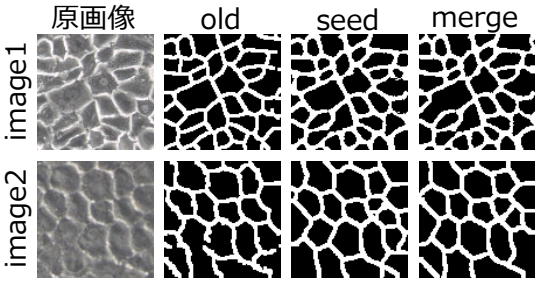


Fig. 10　GP 適用結果（未学習画像）

4.2.2　作成目標画像を用いたフィルタ構築精度の確認

各目標画像を学習画像とし，構築したフィルタから得られた出力画像の一例を Fig. 9 に，未学習画像に対してフィルタ適用した結果の一例を Fig. 10 に示す．また，用いたフィルタは 5 試行中，細胞数の誤差が最も低いフィルタを用いた． Fig. 10，Fig. 9 の結果から，提案手法で作成した目標画像が任意に作成した目標画像と同等の働きをしており，提案手法において作成した目標画像は双方共に，提案システムの学習画像として用いる事が可能であると考えられる．

次に各フィルタを未学習画像に適用して得られた，細胞分割画像の細胞数を計測した結果を Table. 1 に示す．Table. 1 の値は未学習画像の細胞数との誤差率である為，低い程良好な結果となる．Table. 1 の Average とは各画像における誤差率の平均を示す．また，未学習画像の正しい細胞数は目視によって計測を行った．

Table. 1 の結果，対象とする画像によっては任意に作成した目標画像よりも，提案手法によって作成した目標画像を用いた方が良好な結果を示した．特に image2 において良好な結果を示している理由として，任意に作成した目標画像ではドローソフトを用いて分割部位を決定しているのに対して，提案手法ではプロットされた点を用いているとはいえ，基本的に原画像の画素に従って領域を分割している為，良好な分割が出来たのでは無いかと考えられる．また，各画像における誤差率の平均結果からも提案手法が任意に作成するのと同等以上の性能を有している事が確認でき，各画像の結果において merge よりも seed の方が良好な値を示している事から，seeded-watershed を用いる事によって，効果的なフィルタを構築できると考えられる．

Table. 1　未学習画像へのフィルタ適応結果

image/手法	old	seed	merge
Image1	15%	10%	20%
Image2	24%	7%	13%
Image3	11%	4%	7%
Image4	3%	7%	9%
Average	**13%**	**7%**	**12%**

しかしながら，本実験では対象とした画像が 4 枚，GP の試行回数も 5 回と少ない為，対象画像の枚数及び試行回数を増やすことによって，汎用性と安定性の確認を行う必要がある．また，本実験では学習画像として 100 × 100[pixel] の画像を用い画像内の各細胞領域に点をプロットしたが，点の数もしくは学習画像のサイズをどの程度まで確保する必要があるのかを検討する必要がある．

5　まとめ

本稿では，本研究グループが提案している角膜内皮細胞画像の特徴量自動抽出システムにおける学習画像作成の負担を軽減する為に，目標画像作成手法の検討を行った．本手法では，原画像（角膜内皮細胞画像）の細胞領域部分に対して点をプロットし，その点を参考に領域分割手法を用いて領域分割を行う．この領域分割の手法として，watershed+統合法，seeded-watershed の二つの手法を提案し，これらの手法によって作成した画像が本システムの遺伝的プログラミングで用いる学習画像として適当であるかどうかの実験を行った．

実験の結果，両手法において任意に作成した目標画像と同等以上の学習効果を確認でき，seeded-watershed がより有効であることが判明した．今後，本手法の汎用性及び安定性を確認していくと共に，精度の向上を行っていく必要がある．

参考文献

1) N.Koizumi and *et al.* Cultivated corneal endothelialtransplantation in a primate:. *The Journal of Cornea and External Disease, Vol. 27, pp.48-55*, 2011.

2) Carpenter AE and *et al.* image analysis software for identifying and quantifying cell phenotypes. *Genome Biol, 7(10), R100, Oct 31.*, 2006.

3) M. D. Abramoff and *et al.* Image processing with imagej. *Biophotonics international, Volume: 11, Issue: 7 , pp. 36-42*, 2004.

4) 布川将来人, 他. 遺伝的プログラミングを用いた角膜内皮細胞の特徴量自動抽出システムの提案. 第 26 回人工知能学会全国大会, 2012.

5) J.Koza. Genetic programming,on the programming of computers by means of natural selection. *MIT Press*, 1992.

6) D. Hagyard and *et al.* Analysis of watershed algorithms for greyscale images. *Image Processing, 1996. Proceedings., International Conference on volume.3, 41- 44*, 1996.

Federation 型データグリッドにおける データレプリケーションの有効性の検討

藤井 亮助
Ryosuke FUJII

背景：広域に分散した膨大なデータの共有を可能にするデータグリッドにおけるファイル取得遅延問題
研究目的：ファイル取得遅延問題を解決したデータグリッドにおける迅速なファイルの取得の実現
発表の位置づけ：Federation 型データグリッドにおけるデータレプリケーションの有効性の検討
方法：グリッドシミュレーションソフト OptroSim を用いて様々なアクセスパターンのもとでデータレプリケーションの性能評価
結果：総合ジョブ実行時間の平均およびネットワーク効率の点でデータレプリケーションの有効性を確認

1　はじめに

近年，医療機関において膨大なデータが生成され遠隔医療の必要性から複数の病院の医療データの共有が求められている．こうした広域環境におけるデータ共有の基盤には，データグリッドと呼ばれる基盤が使われる．データグリッドとは，広域に分散した資源をネットワークで結びつけ，ひとつの複合したコンピュータシステムとしてデータの保存・利用を可能にするものである．

しかし，データグリッドには，各資源に保存されているファイルサイズが非常に大きく，また各拠点が広域に離れていることからユーザがファイルを取得する際に大幅な遅延とそれに伴うネットワークの帯域幅消費が生じる問題がある．この問題を解決する最も主要な方法としてデータレプリケーションが挙げられる．データレプリケーションとは，ファイルのコピーを複数の拠点に保有して，ユーザが近隣からのファイル取得率を向上させるものである．

そこで本研究では，各拠点において生成された膨大なデータを各拠点で保有した状態で，複数の拠点のデータの共有を可能にさせる Federation 型データグリッドにおける最適なデータレプリケーションの構築を目的とする．

また本稿では，既存のデータレプリケーションアルゴリズムを用いて Federation 型におけるデータレプリケーションの有効性の検討を行う．データレプリケーションの有効性を検討するためにデータグリッドシミュレーションソフト OptorSim を用いて，平均ジョブ実行時間およびネットワーク効率 ENU において性能評価を行った．[1)]

2　データグリッドアーキテクチャ

データグリッドのグリッドアーキテクチャは，Hierachical 型，Federation 型 (Fig. 1)，Hybrid 型，Bottom-Up/Sensor 型の 4 つのモデルに大別される．[2)] その特徴としては，Hierachical 型は単一源のデータ発生からサーバクライアント型の多階層構造を取っている．一方，Federation 型は複数源のデータ発生から各拠点同士が Peer-to-Peer(P2P) 型のアーキテクチャを取っている．

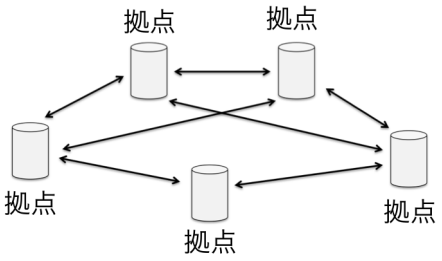


Fig. 1　Federation 型データグリッド

本研究では，各医療機関において生成される膨大な医用画像データを複数の医療機関で共有を可能にするシステムを想定しているため，Federation 型のデータグリッドを用いる．

3　データレプリケーション

データレプリケーションは，データグリッドにおけるデータ取得遅延および帯域幅消費の問題を解決する手法の 1 つである．データレプリケーションには，レプリカ生成の決定，取得レプリカの決定，削除レプリカの決定の 3 つのフェイズが存在する．その中でも特に性能に依存するのは，レプリカ生成の決定である．さらにデータレプリケーションには，レプリカ生成タイミング，レプリカ生成ファイル，およびレプリカ配置場所によって様々な種類のデータレプリケーションが存在する．本稿では，データレプリケーションの有効性を確かめるために，レプリカ生成の決定に関する 2 つの異なるデータレプリケーションアルゴリズム Caching および Economic Model を用いて，複製を行わない NoReplication と比較する．使用するデータレプリケーションアルゴリズムの

詳細は以下の通りである．

3.1 No Replication

この戦略ではファイルの複製を一切行わない．最初のファイルの分散はシミュレーション開始時に決定され，その後実行時間中に変化することはない．No replication は，データレプリケーションの有効性を示すための比較対象として使われる．

3.2 Caching

クライアントがファイルをリクエストした際に，リクエストされたファイルのレプリカをクライアントのローカルノードへ配置する．ローカルストレージにレプリカを配置する十分な容量がない場合はファイル置き換えアルゴリズム Least Recently Used(LRU) により，そのノード内で最も参照の古いファイルを削除した後に新しいレプリカを配置する．

3.3 Economic Model

Economic Model は，Carman らによって提案されたデータレプリケーション戦略であり，最適なレプリケーションの生成に関する決定と削除ファイルの選択に関する決定に利用される．[3] Economic Model の基本的な概念は，将来アクセスされる確率の高いファイルのレプリカを作成し，また新しいレプリカを保存するための容量を確保するために将来的に最もアクセスされる確率が低いファイルを削除ファイルとして選択するものである．そのため，各ファイルに将来アクセスされる確率に基づいて価値を評価する評価関数 E を使用する．また，評価関数 E では，最適な決定を行うために評価対象ファイル f が過去にリクエストされた回数 r の記録をもとに，次の n 回のリクエスト中にアクセスされるファイルの価値を計算する．Economic Model では，ファイル f に関するリクエスト要求を受け取った際に評価関数 E によってファイル f の価値と，その要求を送るローカルノードに存在する全てのファイルの価値を評価関数 E により計算する．もしファイル f が他のどのファイルよりも価値が低ければレプリケーションは行わない．そうでなければ，最も価値の低いファイルが削除ファイルとして選ばれ，またファイル f のレプリケーションが行われる．評価関数 E(f,r,n) は式 (1) によって定義される．

$$E[V(f,k,n),r) = \sum_{i=1}^{n} Pi(f) \tag{1}$$

また，ファイルのリクエストが時間的局性などが高いと仮定し，アクセス履歴が random walk によって表すことができる．random walk は前回アクセスされたファイル ID から近いファイル ID が選択されるアクセスパターンである．このステップが左右対称的な 2 項分布に従うと仮定した場合，将来アクセスされるファイル ID と予想することができる．よって，ファイル f が random walk の i 番目のステップにリクエストされる確率は式 (2) によって定義される．

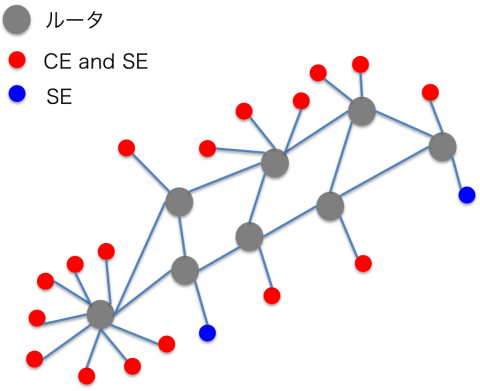


Fig. 2　シミュレーションに用いたグリッドアーキテクチャ

$$Pi(f) = \frac{1}{2^{2iS}} \begin{bmatrix} 2iS \\ f - \overline{f} + iS \end{bmatrix}, |f - \overline{f}| < iS \tag{2}$$

ファイル bar_f は平均ファイル ID，S は 1 ステップあたりに変動する最大幅を表す．また，n は式 (3) に示すように過去にアクセスされた回数 r に依存し，式 (4) より S は r と f によって変動決定される．

$$n = r\frac{T}{T'} \tag{3}$$

$$S = \frac{2}{r}\sum_{j=1}^{r}\frac{(f0 - f_{-j})^2}{j} \tag{4}$$

4 性能評価方法

各データレプリケーションの有効性を評価するためにグリッドシミュレーションソフト OptorSim を用いた．また，有効性の指標として，平均ジョブ実行時間およびネットワーク効率 Effective Network Usage(ENU) の点で比較を行う．本章では，シミュレート時に利用したグリッドアーキテクチャ，各ジョブの詳細について述べる．

4.1 シミュレーション環境

シミュレーションに利用したグリッドアーキテクチャを Fig. 2，また各サイトの容量およびネットワークの性能を Table. 1 に示す．グリッドアーキテクチャは，広域に散在する拠点をルータおよびネットワークにより仮想的に接続されている．また赤で示された拠点はファイルリクエストおよびデータの保有を行う拠点であり，青で

Table. 1　拠点およびネットワークの性能

拠点	Capacity	拠点数
拠点 (赤)	30 GB	17 つ
拠点 (青)	100 GB	2 つ
ネットワーク速度	300MB/s	-

Table. 2　ジョブパラメータ

ジョブ	リクエストファイル数	発生確率
JobA	58 個	50%
JobB	14 個	30%
JobC	12 個	10%
JobD	5 個	4%
JobE	3 個	3%
JobF	6 個	3%

示された拠点はデータ保有のみを行うデータサーバである．ユーザがファイルを取得するジョブを行う際，ユーザがいるローカルの拠点内に目的のファイルがある場合はローカル拠点からファイルを取得する．一方，ローカルの拠点内にない場合は，ネットワークとルータを介して遠方の拠点からファイルを取得する．

本研究におけるジョブは多数のファイルを要求する．本実験で利用するジョブのリクエストファイル数および利用されるジョブの確率を Table. 2 に示す．また，各ファイルのサイズは，最新の 3D 医用画像のファイルサイズを想定して 1 GB を採用した．性能評価で使用したシミュレーションにおけるパラメータを Table. 3 に示す．

シミュレーションの流れは，シミュレーション開始時に全ファイルが Random により各拠点に配置される．その後，指定の時間間隔により各ジョブが指定の確率で呼び出され，各拠点に Random により割り振られる．全ジョブが実行を完了した時点でシミュレーションが終了される．

4.2　アクセスパターン

データレプリケーションの性能は，アクセスパターンによって影響を受ける．そのため，本シュミレーション実験では，Sequeitial，Random，Gaussian random walk，Unitary random walk の 4 つのアクセスパターンを用いて行う．本実験で使用するアクセスパターンのリクエスト例を Fig. 3 に示す．

- Sequential
  ジョブの実行に必要なファイルリストから予め設定された ID に基づいて ID の低い順から連続的にファイルリクエストを行う．ファイルリストの全てのファイルが 1 回ずつアクセスされるため，アクセスパターンには局所性が存在しない．
- Random
  ジョブの実行に必要なファイルリストに対して，ランダムにアクセスを行うファイルを決定する．同一のファイルがアクセスされる場合があるため，Sequential に比べ局所性が強い．
- Gaussian random walk
  前回アクセスを行ったファイル ID を基準としてガウシアン分布に基づいてアクセスを行うファイルを決定する．比較的，同一のファイルがアクセスされる場合が多い．
- Unitary random walk
  前回アクセスを行ったファイル ID から 1 ID 離れたファイルがアクセスされる．ステップする際，基準から高い ID あるいは低い ID の選択は Random によって決定される．また，ファイルリストの最初にアクセスされるファイルはランダムによって決定される．

これらのアクセスパターンの局所性の強さは，Sequential，Random，Gaussian random walk，Unitary random walk の順で強くなる．

4.3　評価指標

本実験では，データレプリケーションの有効性を示すために，平均ジョブ実行時間とネットワーク効率 ENU を性能評価の指標として用いる．ジョブ実行時間は，ジョブが各計算資源にスケジューリングされてからファイル取得などのジョブ実行を経由し，ジョブが完了するまでの時間を表す．一方，ネットワーク効率 ENU はネットワークの利用率を表す指標であり，式 (5) によって定義される．

$$ENU = \frac{RemoteFile\ 取得回数 + Replication\ 回数}{RemoteFile\ 取得回数 + LocalFile\ 取得回数} \quad (5)$$

RemoteFile 取得回数は，ファイルリクエスト時にローカルにない場合に遠方の拠点からリクエストファイルを取得した回数を示す，また，Local ファイル取得回数は，ローカルに存在するファイルを取得した回数を示す．Replication 回数は，レプリケーションを実行した回数を示す．本実験で使用するデータレプリケーションは，Caching および Economic Model ともにリクエストを行ったローカル拠点にレプリカを配置を行う．そのため，ローカルに存在しないファイルをリクエストし，レプリケーション

Table. 3　シミュレーションパラメータ

試行回数	3 回
ジョブ投入回数	200 回
ジョブ投入間隔	10 ms
ファイル配置方式	Random

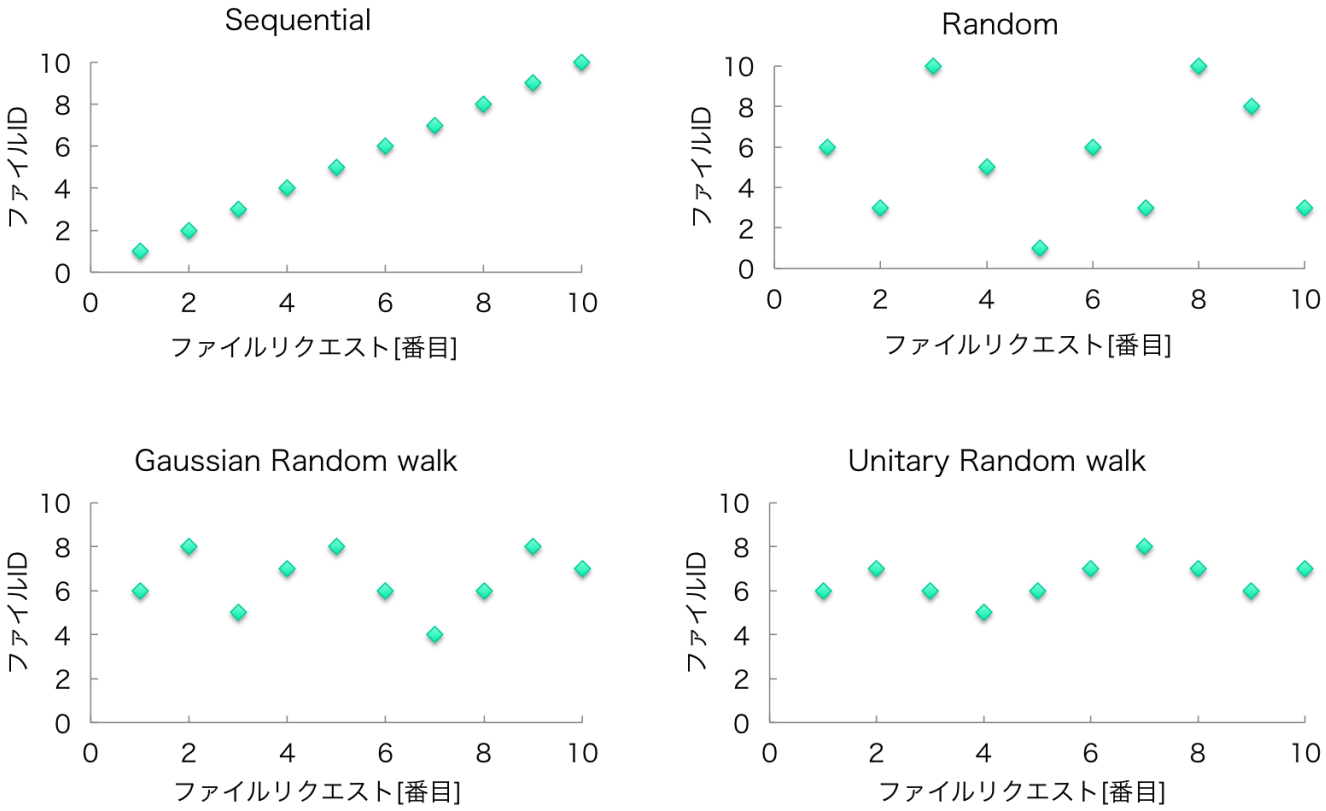


Fig. 3　各アクセスパターンによるリクエストファイルの例

を実行した場合には，Replication 回数が 1 回，RemoteFIle 取得回数が 0 回，LocalFile 取得回数が 1 回が加算される．

ジョブ実行時間は RemoteFile 取得数が減少し LocalFile 取得数が増加するにつれて減少するため，Local アクセスを増加させるレプリケーションによって実現させる．しかし，長期的にシステムを安定的に運用するためには，ネットワークの帯域幅の消費を抑える必要がある．そこで，ジョブ実行時間だけでなく Replication 回数も考慮に入れた ENU という指標を考慮することにより，ファイル取得時間と帯域幅消費の少ない良好なデータレプリケーションを評価できると考えられる．

5　実験結果

5.1　平均ジョブ実行時間

各アクセスパターンにおけるデータレプリケーションの総ジョブ実行時間の平均を Fig. 4 に示す．これは各データレプリケーションによるシミュレーションを 3 回試行した値の平均である．総ジョブ実行時間の平均ではアクセスパターンに関係なく NoReplication に比べて LRU および Economic Model が大幅に実行時間を削減し，データレプリケーションの有効性を示した．また，Caching と EconomicModel では，アクセスパターン Sequential において Economic Model の方が良好が結果を示した．また，これは Hierachical 型データグリッドにおける NoReplication，Caching，および Economic Model の性能比較評価を行っている先行研究と同様の傾向を示す結果となった．[4)]

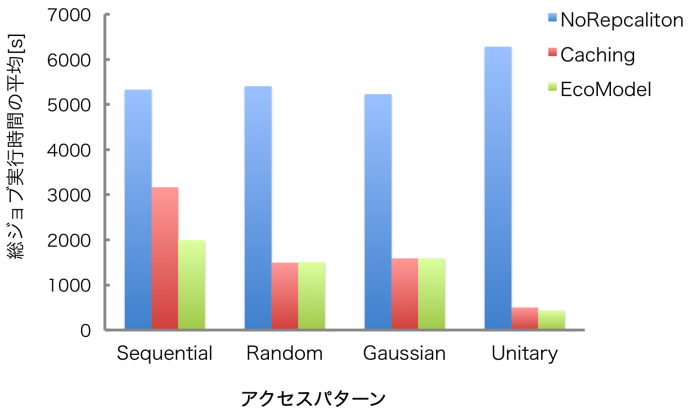


Fig. 4　総ジョブ実行時間の平均

5.2　ネットワーク効率 ENU

各アクセスパターンにおけるデータレプリケーションの ENU を Fig. 5 に示す．データレプリケーションの比較に

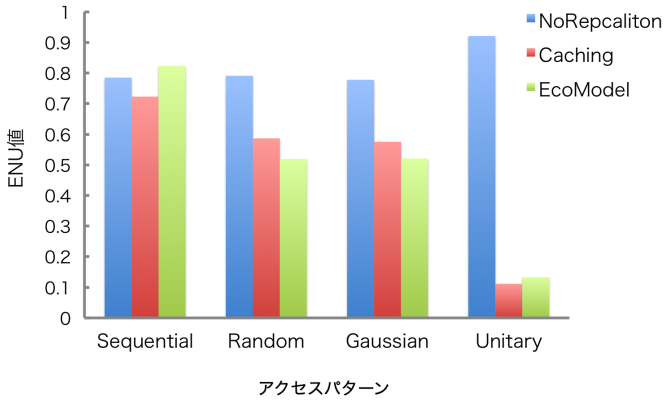


Fig. 5　ネットワーク効率

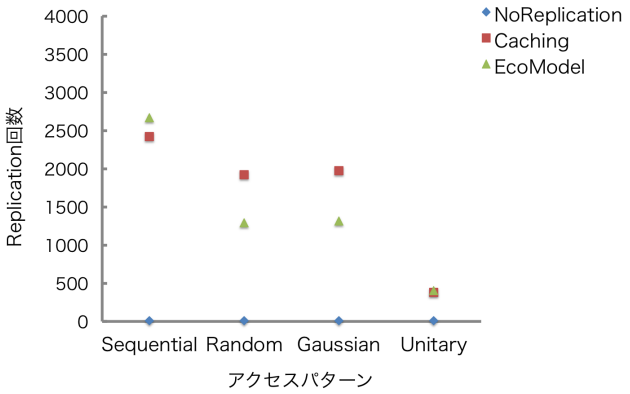


Fig. 6　アクセスパターンの変動における各データレプリケーションの Replication 回数

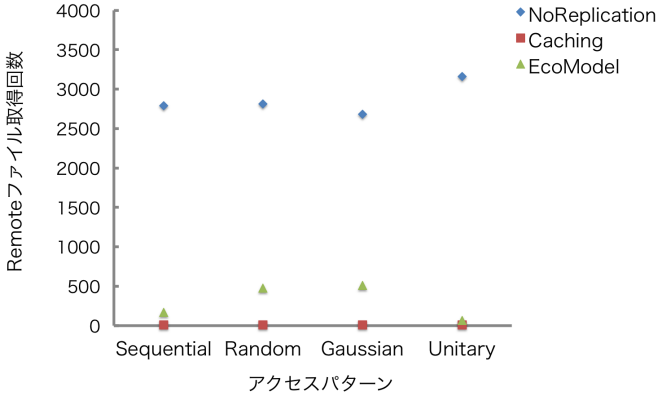


Fig. 7　アクセスパターンの変動における各データレプリケーションの Remote ファイル取得回数

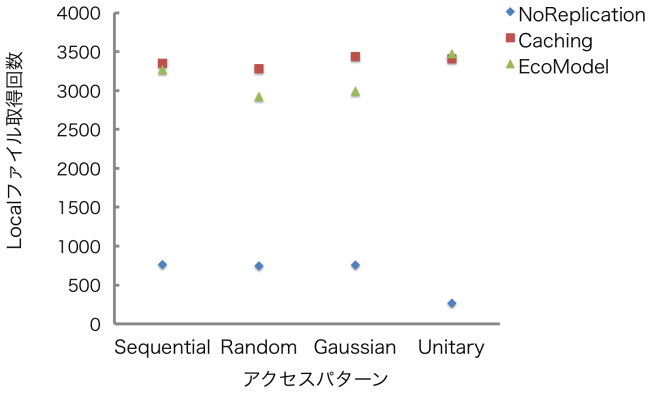


Fig. 8　アクセスパターンの変動における各データレプリケーションの Local ファイル取得回数

おいても Sequential を除く全てのアクセスパターンにおいても NoReplication と比べ Caching および Economic-Model が低い ENU 値を示し，データレプリケーションによりネットワーク効率を向上させることがわかった．Sequential においてのみ，EconomicModel が Caching および NoReplication より高い ENU 値を示し，低いネットワーク効率を示した．

またアクセスパターンの変動では，NoReplicaton がアクセスパターンに関係なく高い ENU を取り，ネットワーク効率の低さを示す結果となった．一方，Caching と EconomicModel では，アクセスパターンが sequential から Unitary へと時間的局所性が増加するにつれて，ENU が減少し高いネットワーク効率を示す結果となった．

6　考察

シミュレーションの結果，レプリケーションによるジョブ実行時間短縮およびネットワーク効率の向上が示された．また，アクセスパターンが Sequential から Random, Gaussian, Unitary へと局所性が強くなるにつれて，データレプリケーションの総ジョブ実行時間の平均およびネットワーク効率が向上する現象が見られた．この理由として，レプリケーションされたファイルがその後のローカルにおいて，有効性に使用されたことが考えられる．Fig. 6 にアクセスパターンの変動における各データレプリケーションの Replication 回数，Fig. 7 にアクセスパターンの変動における各データレプリケーションの Remote ファイル取得回数，Fig. 8 にアクセスパターンの変動における各データレプリケーションの Local ファイル取得回数を示す．

まず，NoReplication の特徴としては，全てのアクセスパターンにおいて Replication 回数が 0 回，Remote ファイル取得回数が 3000 回付近，Local ファイル取得回数が 500 回付近となっている．Caching および Economic Model では，アクセスパターンの局所性が Sequential, Random へと強くなるにつれて Replication 回数が減少している．一方，Local ファイル取得回数は，両アルゴリズムともに一定の値を取り続けている．これは，アクセスパターンにおける局所性が強くなるにつれて同一ファイルへのアクセス率が高くなり，レプリケーションされ

たファイルはその後の繰り返しローカルにおいて使用され，データレプリケーションによる有効性が増大したことを示している．また，ENU が時間的局所性が強いアクセスパターンになるにつれ値が減少したことについても，レプリケーション回数の減少したことが考えられる．

さらに，Economic Model が Caching に比べて ENU が低い値を取ったことについて，Random および Gussian において Economic Model が Caching に比べレプリケーション回数が減少していることから，少ない Replication 回数の中で将来的にアクセスされる確率の高いファイルのレプリケーションに成功していることが原因であると考えられる．

7 まとめ

本稿では，Federation 型データグリッドにおけるデータレプリケーションの有効性を検討した．評価を行ったデータレプリケーションは，Caching および Economic Model であり，比較対象として NoReplication を用いた．様々なアクセスパターンにおいて，データレプリケーションの性能を平均ジョブ実行時間およびネットワーク効率 ENU の 2 点でシミュレーション評価を行った．

その結果，Caching および Economic Model が NoReplication と比べて全てのアクセスパターンにおいて平均ジョブ実行時間を削減することを示した．また ENU においても，Sequential を除く全てのアクセスパターンにおいて，Caching および Economic Model がネットワーク効率を向上させることを示した．特に Economic Model は Caching よりさらに良好な結果を示した．

以上より，Federation 型データグリッドにおいてデータレプリケーションが，システム上の遅延問題を解決する有効な手段であることが確認された．

参考文献

1) J.Ferguson A.P.Millar C.Nicholson K.Stockinger F. Zini D.G.Cameron, R.Schiaffino. OptorSim v2.1 Installation and User Guide. 2006.

2) Srikumar Venugopal, Rajkumar Buyya, and Kotagiri Ramamohanarao. A Taxonomy of Data Grids for Distributed Data Sharing, Management and Processing. *ACM Computing Surveys (CSUR)*, Vol. 38, , 2006.

3) M. Carman, F. Zini, L. Serafini, and K. Stockinger. Towards an Economy-Based Optimisation of File Access and Replication on a Data Grid. *Cluster Computing and the Grid, 2002. 2nd IEEE/ACM International Symposium on*, p. 340, 2002.

4) A. Paul Millar Luigi Capozza Kurt Stockinger William H. Bell, David G. Cameron and Floriano Zini. Optorsim: A Grid Simulator for Studying Dynamic Data Replication Strategies. *International Journal of High Performance Computing Applications*, Vol. 17, pp. 403–416, 2003.